

1. [Preface](#)
2. Unit 1. The Study of Life
  1. [Introduction](#)
  2. [The Science of Biology](#)
  3. [Themes and Concepts of Biology](#)
3. Unit 2. Animal Structure and Function
  1. The Animal Body: Basic Form and Function
    1. [Introduction](#)
    2. [Animal Form and Function](#)
    3. [Animal Primary Tissues](#)
    4. [Homeostasis](#)
  2. Animal Nutrition and the Digestive System
    1. [Introduction](#)
    2. [Digestive Systems](#)
    3. [Nutrition and Energy Production](#)
    4. [Digestive System Processes](#)
    5. [Digestive System Regulation](#)
  3. The Musculoskeletal System
    1. [Introduction](#)
    2. [Types of Skeletal Systems](#)
    3. [Bone](#)
    4. [Joints and Skeletal Movement](#)
    5. [Muscle Contraction and Locomotion](#)
  4. The Nervous System
    1. [Introduction](#)
    2. [Neurons and Glial Cells](#)
    3. [How Neurons Communicate](#)
    4. [The Central Nervous System](#)
    5. [The Peripheral Nervous System](#)
    6. [Nervous System Disorders](#)
  5. Sensory Systems



1. [Introduction](#)
2. [Sensory Processes](#)
3. [Somatosensation](#)
4. [Taste and Smell](#)
5. [Hearing and Vestibular Sensation](#)
6. [Vision](#)
6. The Endocrine System
  1. [Introduction](#)
  2. [Endocrine Glands](#)
  3. [Types of Hormones](#)
  4. [How Hormones Work](#)
  5. [Regulation of Body Processes](#)
  6. [Regulation of Hormone Production](#)
7. The Respiratory System
  1. [Introduction](#)
  2. [Systems of Gas Exchange](#)
  3. [Gas Exchange across Respiratory Surfaces](#)
  4. [Breathing](#)
  5. [Transport of Gases in Human Bodily Fluids](#)
8. The Circulatory System
  1. [Introduction](#)
  2. [Overview of the Circulatory System](#)
  3. [Components of the Blood](#)
  4. [Mammalian Heart and Blood Vessels](#)
  5. [Blood Flow and Blood Pressure Regulation](#)
9. Osmotic Regulation and Excretion
  1. [Introduction](#)
  2. [Osmoregulation and Osmotic Balance](#)
  3. [The Kidneys and Osmoregulatory Organs](#)
  4. [Excretion Systems](#)
  5. [Nitrogenous Wastes](#)
  6. [Hormonal Control of Osmoregulatory Functions](#)

## 10. The Immune System

1. [Introduction](#)
2. [Innate Immune Response](#)
3. [Adaptive Immune Response](#)
4. [Antibodies](#)
5. [Disruptions in the Immune System](#)

## 11. Animal Reproduction and Development

1. [Introduction](#)
2. [Reproduction Methods](#)
3. [Fertilization](#)
4. [Human Reproductive Anatomy and Gametogenesis](#)
5. [Hormonal Control of Human Reproduction](#)
6. [Human Pregnancy and Birth](#)
7. [Fertilization and Early Embryonic Development](#)
8. [Organogenesis and Vertebrate Formation](#)

## 4. Unit 3. Evolutionary Processes

### 1. Evolution and the Origin of Species

1. [Introduction](#)
2. [Understanding Evolution](#)
3. [Formation of New Species](#)
4. [Reconnection and Rates of Speciation](#)

### 2. The Evolution of Populations

1. [Introduction](#)
2. [Population Evolution](#)
3. [Population Genetics](#)
4. [Adaptive Evolution](#)

### 3. Phylogenies and the History of Life

1. [Introduction](#)
2. [Organizing Life on Earth](#)
3. [Determining Evolutionary Relationships](#)
4. [Perspectives on the Phylogenetic Tree](#)

## 5. Unit 3. Introduction to the Plant Kingdom

## 1. Plant Diversity

### 1. Seedless Plants

1. [Introduction](#)
2. [Early Plant Life](#)
3. [Green Algae: Precursors of Land Plants](#)
4. [Bryophytes](#)
5. [Seedless Vascular Plants](#)

### 2. Seed Plants

1. [Introduction](#)
2. [Evolution of Seed Plants](#)
3. [Gymnosperms](#)
4. [Angiosperms](#)
5. [The Role of Seed Plants](#)

## 6. Unit 4. Plant Structure and Function

### 1. Plant Form and Physiology

1. [Introduction](#)
2. [The Plant Body](#)
3. [Stems](#)
4. [Roots](#)
5. [Leaves](#)
6. [Transport of Water and Solutes in Plants](#)
7. [Plant Sensory Systems and Responses](#)

### 2. Soil and Plant Nutrition

1. [Introduction](#)
2. [Nutritional Requirements of Plants](#)
3. [The Soil](#)
4. [Nutritional Adaptations of Plants](#)

### 3. Plant Reproduction

1. [Introduction](#)
2. [Reproductive Development and Structure](#)
3. [Pollination and Fertilization](#)
4. [Asexual Reproduction](#)

## 7. Unit 5. Ecology

### 1. Ecology and the Biosphere

1. [Introduction](#)
2. [The Scope of Ecology](#)
3. [Biogeography](#)
4. [Terrestrial Biomes](#)
5. [Aquatic Biomes](#)
6. [Climate and the Effects of Global Climate Change](#)

### 2. Population and Community Ecology

1. [Introduction](#)
2. [Population Demography](#)
3. [Life Histories and Natural Selection](#)
4. [Environmental Limits to Population Growth](#)
5. [Population Dynamics and Regulation](#)
6. [Human Population Growth](#)
7. [Community Ecology](#)
8. [Behavioral Biology: Proximate and Ultimate Causes of Behavior](#)

### 3. Ecosystems

1. [Introduction](#)
2. [Ecology of Ecosystems](#)
3. [Energy Flow through Ecosystems](#)
4. [Biogeochemical Cycles](#)

### 4. Conservation Biology and Biodiversity

1. [Introduction](#)
2. [The Biodiversity Crisis](#)
3. [The Importance of Biodiversity to Human Life](#)
4. [Threats to Biodiversity](#)
5. [Preserving Biodiversity](#)

## 8. [The Periodic Table of Elements](#)

## Preface

Biology is designed for multi-semester biology courses for science majors. It is grounded on an evolutionary basis and includes exciting features that highlight careers in the biological sciences and everyday applications of the concepts at hand. To meet the needs of today's instructors and students, some content has been strategically condensed while maintaining the overall scope and coverage of traditional texts for this course. Instructors can customize the book, adapting it to the approach that works best in their classroom. Biology also includes an innovative art program that incorporates critical thinking and clicker questions to help students understand—and apply—key concepts.

Welcome to *Biology*, an OpenStax resource. This textbook was written to increase student access to high-quality learning materials, maintaining highest standards of academic rigor at little to no cost.

## About OpenStax

OpenStax is a nonprofit based at Rice University, and it's our mission to improve student access to education. Our first openly licensed college textbook was published in 2012, and our library has since scaled to over 20 books for college and AP courses used by hundreds of thousands of students. Our adaptive learning technology, designed to improve learning outcomes through personalized educational paths, is being piloted in college courses throughout the country. Through our partnerships with philanthropic foundations and our alliance with other educational resource organizations, OpenStax is breaking down the most common barriers to learning and empowering students and instructors to succeed.

## About OpenStax's Resources

### Customization

*Biology* is licensed under a Creative Commons Attribution 4.0 International (CC BY) license, which means that you can distribute, remix, and build

upon the content, as long as you provide attribution to OpenStax and its content contributors.

Because our books are openly licensed, you are free to use the entire book or pick and choose the sections that are most relevant to the needs of your course. Feel free to remix the content by assigning your students certain chapters and sections in your syllabus, in the order that you prefer. You can even provide a direct link in your syllabus to the sections in the web view of your book.

Instructors also have the option of creating a customized version of their OpenStax book. The custom version can be made available to students in low-cost print or digital form through their campus bookstore. Visit your book page on [openstax.org](https://openstax.org) for more information.

## **Errata**

All OpenStax textbooks undergo a rigorous review process. However, like any professional-grade textbook, errors sometimes occur. Since our books are web based, we can make updates periodically when deemed pedagogically necessary. If you have a correction to suggest, submit it through the link on your book page on [openstax.org](https://openstax.org). Subject matter experts review all errata suggestions. OpenStax is committed to remaining transparent about all updates, so you will also find a list of past errata changes on your book page on [openstax.org](https://openstax.org).

## **Format**

You can access this textbook for free in web view or PDF through [openstax.org](https://openstax.org), and in low-cost print and iBooks editions.

## **About Biology**

*Biology* is designed to cover the scope and sequence requirements of a typical two-semester biology course for science majors. The text provides comprehensive coverage of foundational research and core biology concepts through an evolutionary lens. *Biology* includes rich features that engage students in scientific inquiry, highlight careers in the biological sciences, and offer everyday applications. The book also includes clicker questions to help students understand—and apply—key concepts.

## Coverage and Scope

In developing *Biology*, we listened to hundreds of General Biology instructors who readily provided feedback about their courses, students, challenges, and hopes for innovation. The expense of textbooks and related items did prove to be a barrier to learning. But more importantly, these teachers suggested improvements for the textbook, which would ultimately lead to more meaningful and memorable learning experiences for students.

The result is a book that addresses a core organizational reality of the course and its materials—the sheer breadth of the topical coverage. We provide a thorough treatment of biology’s foundational concepts while condensing selected topics in response to the market’s request for a textbook with a scope that is manageable for instructors and students alike. We also strive to make biology, as a discipline, interesting and accessible to students. In addition to a comprehensive coverage of core concepts and foundational research, we have incorporated features that draw learners into the discipline in meaningful ways.

The pedagogical choices, chapter arrangements, and learning objective fulfillment were developed and vetted with the feedback of another one hundred reviewers, who thoroughly read the material and offered detailed critical commentary.

**Unit 1: The Chemistry of Life.** Our opening unit introduces students to the sciences, including the scientific method and the fundamental concepts of chemistry and physics that provide a framework within which learners comprehend biological processes.

Unit 2: **The Cell**. Students will gain solid understanding of the structures, functions, and processes of the most basic unit of life: the cell.

Unit 3: **Genetics**. Our comprehensive genetics unit takes learners from the earliest experiments that revealed the basis of genetics through the intricacies of DNA to current applications in the emerging studies of biotechnology and genomics.

Unit 4: **Evolutionary Processes**. The core concepts of evolution are discussed in this unit with examples illustrating evolutionary processes. Additionally, the evolutionary basis of biology reappears throughout the textbook in general discussion and is reinforced through special call-out features highlighting specific evolution-based topics.

Unit 5: **Biological Diversity**. The diversity of life is explored with detailed study of various organisms and discussion of emerging phylogenetic relationships. This unit moves from viruses to living organisms like bacteria, discusses the organisms formerly grouped as protists, and devotes multiple chapters to plant and animal life.

Unit 6: **Plant Structure and Function**. Our plant unit thoroughly covers the fundamental knowledge of plant life essential to an introductory biology course.

Unit 7: **Animal Structure and Function**. An introduction to the form and function of the animal body is followed by chapters on specific body systems and processes. This unit touches on the biology of all organisms while maintaining an engaging focus on human anatomy and physiology that helps students connect to the topics.

Unit 8: **Ecology**. Ecological concepts are broadly covered in this unit, with features highlighting localized, real-world issues of conservation and biodiversity.

## **Pedagogical Foundation and Features**

*Biology* is grounded in a solid scientific base, with features that engage the students in scientific inquiry, including:



**Evolution Connection** features uphold the importance of evolution to all biological study through discussions like “The Evolution of Metabolic Pathways” and “Algae and Evolutionary Paths to Photosynthesis.”

**Scientific Method Connection** call-outs walk students through actual or thought experiments that elucidate the steps of the scientific process as applied to the topic. Features include “Determining the Time Spent in Cell Cycle Stages” and “Testing the Hypothesis of Independent Assortment.”

**Career Connection** features present information on a variety of careers in the biological sciences, introducing students to the educational requirements and day-to-day work life of a variety of professions, such as microbiologist, ecologist, neurologist, and forensic scientist.

**Everyday Connection** features tie biological concepts to emerging issues and discuss science in terms of everyday life. Topics include “Chesapeake Bay” and “Can Snail Venom Be Used as a Pharmacological Pain Killer?”

## **Art and Animations That Engage**

Our art program takes a straightforward approach designed to help students learn the concepts of biology through simple, effective illustrations, photos, and micrographs. *Biology* also incorporates links to relevant animations and interactive exercises that help bring biology to life for students.

**Art Connection** features call out core figures in each chapter for student study. Questions about key figures, including clicker questions that can be used in the classroom, engage students’ critical thinking to ensure genuine understanding.

**Link to Learning** features direct students to online interactive exercises and animations to add a fuller context to core content.

## **Additional Resources**

## **Student and Instructor Resources**

We've compiled additional resources for both students and instructors, including Getting Started Guides, an instructor solution manual, supplemental test items, and PowerPoint slides. Instructor resources require a verified instructor account, which can be requested on your [openstax.org](https://openstax.org) log-in. Take advantage of these resources to supplement your OpenStax book.

## **Partner Resources**

OpenStax Partners are our allies in the mission to make high-quality learning materials affordable and accessible to students and instructors everywhere. Their tools integrate seamlessly with our OpenStax titles at a low cost. To access the partner resources for your text, visit your book page on [openstax.org](https://openstax.org).

## **About the Authors**

### **Senior Contributing Authors**

Yael Avissar (Cell Biology), Rhode Island College

Jung Choi (Genetics), Georgia Institute of Technology

Jean DeSaix (Evolution), University of North Carolina at Chapel Hill

Vladimir Jurukovski (Animal Physiology), Suffolk County Community College

Robert Wise (Plant Biology), University of Wisconsin, Oshkosh

Connie Rye (General Content Lead), East Mississippi Community College

### **Contributing Authors and Reviewers**

Julie Adams, Aurora University

Summer Allen, Brown University

James Bader, Case Western Reserve University

David Bailey, St. Norbert College

Mark Belk, Brigham Young University

Nancy Boury, Iowa State University

Lisa Bonneau, Metropolitan Community College – Blue River  
Graciela Brelles-Marino, California State University Pomona  
Mark Browning, Purdue University  
Sue Chaplin, University of St. Thomas  
George Cline, Jacksonville State University  
Deb Cook, Georgia Gwinnett College  
Diane Day, Clayton State University  
Frank Dirrigl, The University of Texas Pan American  
Waneene Dorsey, Grambling State University  
Nick Downey, University of Wisconsin La Crosse  
Rick Duhrkopf, Baylor University  
Kristy Duran, Adams State University  
Stan Eisen, Christian Brothers University  
Brent Ewers, University of Wyoming  
Myriam Feldman, Lake Washington Institute of Technology  
Michael Fine, Virginia Commonwealth University  
Linda Flora, Delaware County Community College  
Thomas Freeland, Walsh University  
David Gris , Texas A & M University – Corpus Christi  
Andrea Hazard, SUNY Cortland  
Michael Hedrick, University of North Texas  
Linda Hensel, Mercer University  
Mark Kopeny, University of Virginia  
Norman Johnson, University of Massachusetts Amherst  
Grace Lasker, Lake Washington Institute of Technology; Walden University  
Sandy Latourelle, SUNY Plattsburgh  
Theo Light, Shippensburg University  
Clark Lindgren, Grinnell College  
James Malcolm, University of Redlands  
Mark Meade, Jacksonville State University  
Richard Merritt, Houston Community College  
James Mickle, North Carolina State University  
Jasleen Mishra, Houston Community College  
Dudley Moon, Albany College of Pharmacy and Health Sciences  
Shobhana Natarajan, Brookhaven College  
Jonas Okeagu, Fayetteville State University  
Diana Oliveras, University of Colorado Boulder

John Peters, College of Charleston  
Joel Piperberg, Millersville University  
Johanna Porter-Kelley, Winston-Salem State University  
Robyn Puffenbarger, Bridgewater College  
Dennis Revie, California Lutheran University  
Ann Rushing, Baylor University  
Sangha Saha, City College of Chicago  
Edward Saiff, Ramapo College of New Jersey  
Brian Shmaefsky, Lone Star College System  
Robert Sizemore, Alcorn State University  
Marc Smith, Sinclair Community College  
Frederick Spiegel, University of Arkansas  
Frederick Sproull, La Roche College  
Bob Sullivan, Marist College  
Mark Sutherland, Hendrix College  
Toure Thompson, Alabama A&M University  
Scott Thomson, University of Wisconsin – Parkside  
Allison van de Meene, University of Melbourne  
Mary White, Southeastern Louisiana University  
Steven Wilt, Bellarmine University  
James Wise, Hampton University  
Renna Wolfe  
Virginia Young, Mercer University  
Leslie Zeman, University of Washington  
Daniel Zurek, Pittsburg State University  
Shobhana Natarajan, Alcon Laboratories, Inc.

## Introduction

class="introduction"

This NASA image is a composite of several satellite-based views of Earth. To make the whole-Earth image, NASA scientists combine observations of different parts of the planet. (credit: NASA/GSFC/NOAA/USGS)

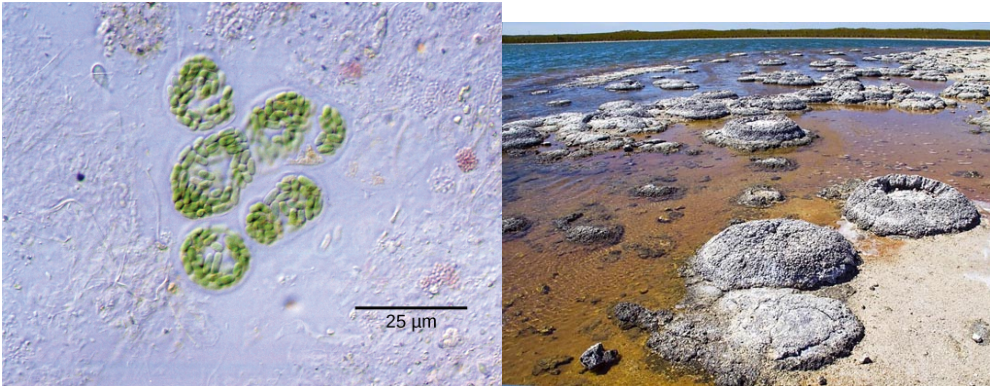


Viewed from space, Earth offers no clues about the diversity of life forms that reside there. The first forms of life on Earth are thought to have been microorganisms that existed for billions of years in the ocean before plants and animals appeared. The mammals, birds, and flowers so familiar to us are all relatively recent, originating 130 to 200 million years ago. Humans have inhabited this planet for only the last 2.5 million years, and only in the last 200,000 years have humans started looking like we do today.

## The Science of Biology

By the end of this section, you will be able to:

- Identify the shared characteristics of the natural sciences
- Summarize the steps of the scientific method
- Compare inductive reasoning with deductive reasoning
- Describe the goals of basic science and applied science

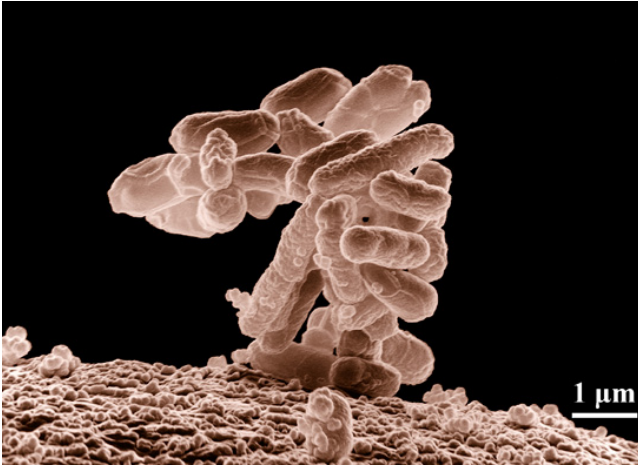


Formerly called blue-green algae, these (a) cyanobacteria, shown here at 300x magnification under a light microscope, are some of Earth's oldest life forms.

These (b) stromatolites along the shores of Lake Thetis in Western Australia are ancient structures formed by the layering of cyanobacteria in shallow waters. (credit a: modification of work by NASA; credit b: modification of work by Ruth Ellison; scale-bar data from Matt Russell)

What is biology? In simple terms, **biology** is the study of living organisms and their interactions with one another and their environments. This is a very broad definition because the scope of biology is vast. Biologists may study anything from the microscopic or submicroscopic view of a cell to ecosystems and the whole living planet ([\[link\]](#)). Listening to the daily news, you will quickly realize how many aspects of biology are discussed every day. For example, recent news topics include *Escherichia coli* ([\[link\]](#)) outbreaks in spinach and *Salmonella* contamination in peanut butter. Other subjects include efforts toward finding a cure for AIDS, Alzheimer's

disease, and cancer. On a global scale, many researchers are committed to finding ways to protect the planet, solve environmental issues, and reduce the effects of climate change. All of these diverse endeavors are related to different facets of the discipline of biology.



*Escherichia coli* (*E. coli*) bacteria, seen in this scanning electron micrograph, are normal residents of our digestive tracts that aid in the absorption of vitamin K and other nutrients. However, virulent strains are sometimes responsible for disease outbreaks. (credit: Eric Erbe, digital colorization by Christopher Pooley, both of USDA, ARS, EMU)

## The Process of Science

Biology is a science, but what exactly is science? What does the study of biology share with other scientific disciplines? **Science** (from the Latin *scientia*, meaning “knowledge”) can be defined as knowledge that covers general truths or the operation of general laws, especially when acquired and tested by the scientific method. It becomes clear from this definition

that the application of the scientific method plays a major role in science. The **scientific method** is a method of research with defined steps that include experiments and careful observation.

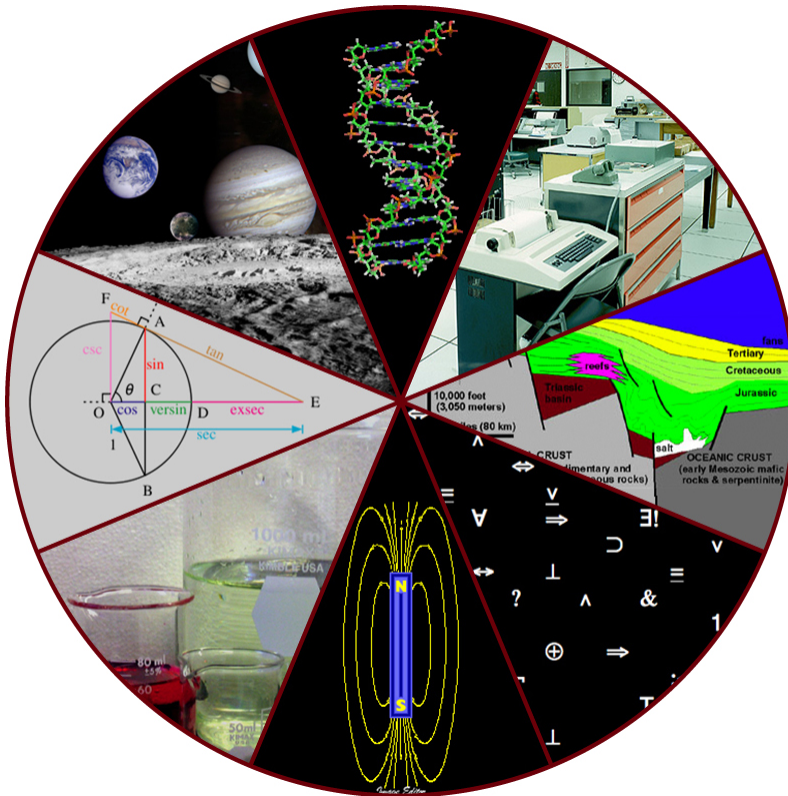
The steps of the scientific method will be examined in detail later, but one of the most important aspects of this method is the testing of hypotheses by means of repeatable experiments. A **hypothesis** is a suggested explanation for an event, which can be tested. Although using the scientific method is inherent to science, it is inadequate in determining what science is. This is because it is relatively easy to apply the scientific method to disciplines such as physics and chemistry, but when it comes to disciplines like archaeology, psychology, and geology, the scientific method becomes less applicable as it becomes more difficult to repeat experiments.

These areas of study are still sciences, however. Consider archeology—even though one cannot perform repeatable experiments, hypotheses may still be supported. For instance, an archeologist can hypothesize that an ancient culture existed based on finding a piece of pottery. Further hypotheses could be made about various characteristics of this culture, and these hypotheses may be found to be correct or false through continued support or contradictions from other findings. A hypothesis may become a verified theory. A **theory** is a tested and confirmed explanation for observations or phenomena. Science may be better defined as fields of study that attempt to comprehend the nature of the universe.

## Natural Sciences

What would you expect to see in a museum of natural sciences? Frogs? Plants? Dinosaur skeletons? Exhibits about how the brain functions? A planetarium? Gems and minerals? Or, maybe all of the above? Science includes such diverse fields as astronomy, biology, computer sciences, geology, logic, physics, chemistry, and mathematics ([\[link\]](#)). However, those fields of science related to the physical world and its phenomena and processes are considered **natural sciences**. Thus, a museum of natural sciences might contain any of the items listed above.





The diversity of scientific fields includes astronomy, biology, computer science, geology, logic, physics, chemistry, mathematics, and many other fields. (credit: “Image Editor”/Flickr)

There is no complete agreement when it comes to defining what the natural sciences include, however. For some experts, the natural sciences are astronomy, biology, chemistry, earth science, and physics. Other scholars choose to divide natural sciences into **life sciences**, which study living things and include biology, and **physical sciences**, which study nonliving matter and include astronomy, geology, physics, and chemistry. Some disciplines such as biophysics and biochemistry build on both life and physical sciences and are interdisciplinary. Natural sciences are sometimes referred to as “hard science” because they rely on the use of quantitative data; social sciences that study society and human behavior are more likely to use qualitative assessments to drive investigations and findings.

Not surprisingly, the natural science of biology has many branches or subdisciplines. Cell biologists study cell structure and function, while biologists who study anatomy investigate the structure of an entire organism. Those biologists studying physiology, however, focus on the internal functioning of an organism. Some areas of biology focus on only particular types of living things. For example, botanists explore plants, while zoologists specialize in animals.

## **Scientific Reasoning**

One thing is common to all forms of science: an ultimate goal “to know.” Curiosity and inquiry are the driving forces for the development of science. Scientists seek to understand the world and the way it operates. To do this, they use two methods of logical thinking: inductive reasoning and deductive reasoning.

**Inductive reasoning** is a form of logical thinking that uses related observations to arrive at a general conclusion. This type of reasoning is common in descriptive science. A life scientist such as a biologist makes observations and records them. These data can be qualitative or quantitative, and the raw data can be supplemented with drawings, pictures, photos, or videos. From many observations, the scientist can infer conclusions (inductions) based on evidence. Inductive reasoning involves formulating generalizations inferred from careful observation and the analysis of a large amount of data. Brain studies provide an example. In this type of research, many live brains are observed while people are doing a specific activity, such as viewing images of food. The part of the brain that “lights up” during this activity is then predicted to be the part controlling the response to the selected stimulus, in this case, images of food. The “lighting up” of the various areas of the brain is caused by excess absorption of radioactive sugar derivatives by active areas of the brain. The resultant increase in radioactivity is observed by a scanner. Then, researchers can stimulate that part of the brain to see if similar responses result.

Deductive reasoning or deduction is the type of logic used in hypothesis-based science. In deductive reason, the pattern of thinking moves in the opposite direction as compared to inductive reasoning. **Deductive reasoning** is a form of logical thinking that uses a general principle or law to forecast specific results. From those general principles, a scientist can extrapolate and predict the specific results that would be valid as long as the general principles are valid. Studies in climate change can illustrate this type of reasoning. For example, scientists may predict that if the climate becomes warmer in a particular region, then the distribution of plants and animals should change. These predictions have been made and tested, and many such changes have been found, such as the modification of arable areas for agriculture, with change based on temperature averages.

Both types of logical thinking are related to the two main pathways of scientific study: descriptive science and hypothesis-based science. **Descriptive (or discovery) science**, which is usually inductive, aims to observe, explore, and discover, while **hypothesis-based science**, which is usually deductive, begins with a specific question or problem and a potential answer or solution that can be tested. The boundary between these two forms of study is often blurred, and most scientific endeavors combine both approaches. The fuzzy boundary becomes apparent when thinking about how easily observation can lead to specific questions. For example, a gentleman in the 1940s observed that the burr seeds that stuck to his clothes and his dog's fur had a tiny hook structure. On closer inspection, he discovered that the burrs' gripping device was more reliable than a zipper. He eventually developed a company and produced the hook-and-loop fastener popularly known today as Velcro. Descriptive science and hypothesis-based science are in continuous dialogue.

## The Scientific Method

Biologists study the living world by posing questions about it and seeking science-based responses. This approach is common to other sciences as well and is often referred to as the scientific method. The scientific method was used even in ancient times, but it was first documented by England's Sir Francis Bacon (1561–1626) ([\[link\]](#)), who set up inductive methods for scientific inquiry. The scientific method is not exclusively used by

biologists but can be applied to almost all fields of study as a logical, rational problem-solving method.



Sir Francis Bacon  
(1561–1626) is  
credited with being  
the first to define  
the scientific  
method. (credit:  
Paul van Somer)

The scientific process typically starts with an observation (often a problem to be solved) that leads to a question. Let's think about a simple problem that starts with an observation and apply the scientific method to solve the problem. One Monday morning, a student arrives at class and quickly discovers that the classroom is too warm. That is an observation that also describes a problem: the classroom is too warm. The student then asks a question: "Why is the classroom so warm?"

## Proposing a Hypothesis

Recall that a hypothesis is a suggested explanation that can be tested. To solve a problem, several hypotheses may be proposed. For example, one hypothesis might be, “The classroom is warm because no one turned on the air conditioning.” But there could be other responses to the question, and therefore other hypotheses may be proposed. A second hypothesis might be, “The classroom is warm because there is a power failure, and so the air conditioning doesn’t work.”

Once a hypothesis has been selected, the student can make a prediction. A prediction is similar to a hypothesis but it typically has the format “If . . . then . . . .” For example, the prediction for the first hypothesis might be, “*If* the student turns on the air conditioning, *then* the classroom will no longer be too warm.”

## Testing a Hypothesis

A valid hypothesis must be testable. It should also be **falsifiable**, meaning that it can be disproven by experimental results. Importantly, science does not claim to “prove” anything because scientific understandings are always subject to modification with further information. This step—openness to disproving ideas—is what distinguishes sciences from non-sciences. The presence of the supernatural, for instance, is neither testable nor falsifiable. To test a hypothesis, a researcher will conduct one or more experiments designed to eliminate one or more of the hypotheses. Each experiment will have one or more variables and one or more controls. A **variable** is any part of the experiment that can vary or change during the experiment. The **control group** contains every feature of the experimental group except it is not given the manipulation that is hypothesized about. Therefore, if the results of the experimental group differ from the control group, the difference must be due to the hypothesized manipulation, rather than some outside factor. Look for the variables and controls in the examples that follow. To test the first hypothesis, the student would find out if the air conditioning is on. If the air conditioning is turned on but does not work, there should be another reason, and this hypothesis should be rejected. To

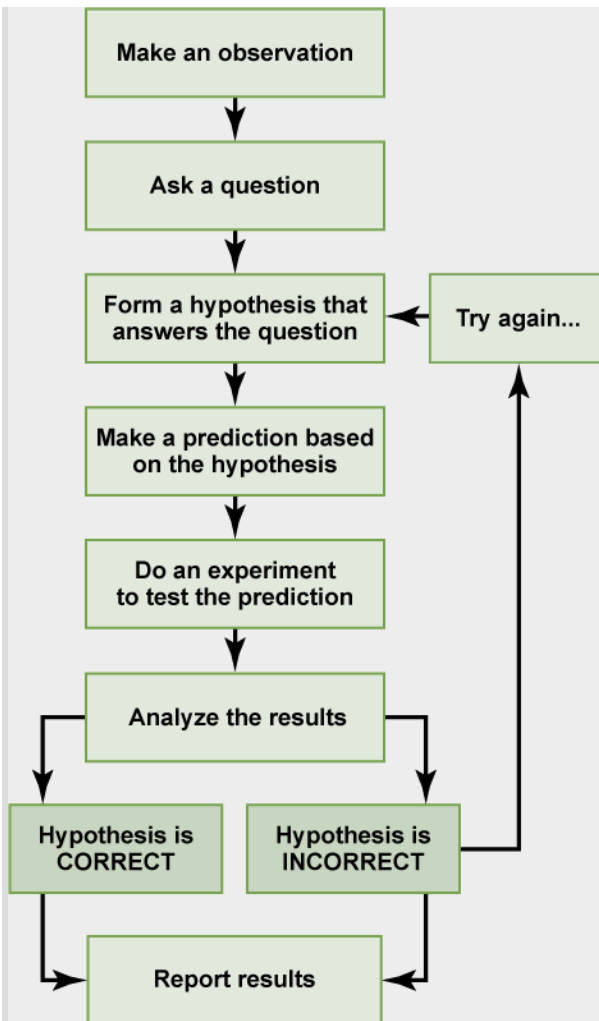
test the second hypothesis, the student could check if the lights in the classroom are functional. If so, there is no power failure and this hypothesis should be rejected. Each hypothesis should be tested by carrying out appropriate experiments. Be aware that rejecting one hypothesis does not determine whether or not the other hypotheses can be accepted; it simply eliminates one hypothesis that is not valid ([\[link\]](#)). Using the scientific method, the hypotheses that are inconsistent with experimental data are rejected.

While this “warm classroom” example is based on observational results, other hypotheses and experiments might have clearer controls. For instance, a student might attend class on Monday and realize she had difficulty concentrating on the lecture. One observation to explain this occurrence might be, “When I eat breakfast before class, I am better able to pay attention.” The student could then design an experiment with a control to test this hypothesis.

In hypothesis-based science, specific results are predicted from a general premise. This type of reasoning is called deductive reasoning: deduction proceeds from the general to the particular. But the reverse of the process is also possible: sometimes, scientists reach a general conclusion from a number of specific observations. This type of reasoning is called inductive reasoning, and it proceeds from the particular to the general. Inductive and deductive reasoning are often used in tandem to advance scientific knowledge ([\[link\]](#)). In recent years a new approach of testing hypotheses has developed as a result of an exponential growth of data deposited in various databases. Using computer algorithms and statistical analyses of data in databases, a new field of so-called “data research” (also referred to as “in silico” research) provides new methods of data analyses and their interpretation. This will increase the demand for specialists in both biology and computer science, a promising career opportunity.

**Note:**

Art Connection



The scientific method consists of a series of well-defined steps. If a hypothesis is not supported by experimental data, a new hypothesis can be proposed.

In the example below, the scientific method is used to solve an everyday problem. Order the scientific method steps (numbered items) with the process of solving the everyday problem (lettered items). Based on the results of the experiment, is the hypothesis correct? If it is incorrect, propose some alternative hypotheses.

1. Observation

2. Question
3. Hypothesis (answer)
4. Prediction
5. Experiment
6. Result

- a. There is something wrong with the electrical outlet.
- b. If something is wrong with the outlet, my coffeemaker also won't work when plugged into it.
- c. My toaster doesn't toast my bread.
- d. I plug my coffee maker into the outlet.
- e. My coffeemaker works.
- f. Why doesn't my toaster work?

## Note:

### Art Connection

Two Types of Reasoning	
<b>Inductive reasoning:</b> from a number of observations, a general conclusion is drawn.	<b>Deductive reasoning:</b> from a general premise, specific results are predicted.
<b>Observations</b>	<b>General premise</b>
<ul style="list-style-type: none"> <li>Members of a species are not all the same.</li> <li>Individuals compete for resources.</li> <li>Species are generally adapted to their environment.</li> </ul>	Individuals most adapted to their environment are more likely to survive and pass their traits on to the next generation.
↓	↓
<b>Conclusion</b>	<b>Predicted results</b>
Individuals most adapted to their environment are more likely to survive and pass their traits to the next generation.	If the average temperature in an ecosystem increases due to climate change, individuals better adapted to warmer temperatures will outcompete those that are not.



Scientists use two types of reasoning, inductive and deductive reasoning, to advance scientific knowledge. As is the case in this example, the conclusion from inductive reasoning can often become the premise for inductive reasoning.

Decide if each of the following is an example of inductive or deductive reasoning.

1. All flying birds and insects have wings. Birds and insects flap their wings as they move through the air. Therefore, wings enable flight.
2. Insects generally survive mild winters better than harsh ones. Therefore, insect pests will become more problematic if global temperatures increase.
3. Chromosomes, the carriers of DNA, separate into daughter cells during cell division. Therefore, DNA is the genetic material.
4. Animals as diverse as humans, insects, and wolves all exhibit social behavior. Therefore, social behavior must have an evolutionary advantage.

The scientific method may seem too rigid and structured. It is important to keep in mind that, although scientists often follow this sequence, there is flexibility. Sometimes an experiment leads to conclusions that favor a change in approach; often, an experiment brings entirely new scientific questions to the puzzle. Many times, science does not operate in a linear fashion; instead, scientists continually draw inferences and make generalizations, finding patterns as their research proceeds. Scientific reasoning is more complex than the scientific method alone suggests. Notice, too, that the scientific method can be applied to solving problems that aren't necessarily scientific in nature.

## Two Types of Science: Basic Science and Applied Science

The scientific community has been debating for the last few decades about the value of different types of science. Is it valuable to pursue science for the sake of simply gaining knowledge, or does scientific knowledge only have worth if we can apply it to solving a specific problem or to bettering our lives? This question focuses on the differences between two types of science: basic science and applied science.

**Basic science** or “pure” science seeks to expand knowledge regardless of the short-term application of that knowledge. It is not focused on developing a product or a service of immediate public or commercial value. The immediate goal of basic science is knowledge for knowledge’s sake, though this does not mean that, in the end, it may not result in a practical application.

In contrast, **applied science** or “technology,” aims to use science to solve real-world problems, making it possible, for example, to improve a crop yield, find a cure for a particular disease, or save animals threatened by a natural disaster ([\[link\]](#)). In applied science, the problem is usually defined for the researcher.



After Hurricane Ike struck the Gulf Coast in 2008, the U.S. Fish and Wildlife Service rescued this

brown pelican. Thanks to applied science, scientists knew how to rehabilitate the bird. (credit: FEMA)

Some individuals may perceive applied science as “useful” and basic science as “useless.” A question these people might pose to a scientist advocating knowledge acquisition would be, “What for?” A careful look at the history of science, however, reveals that basic knowledge has resulted in many remarkable applications of great value. Many scientists think that a basic understanding of science is necessary before an application is developed; therefore, applied science relies on the results generated through basic science. Other scientists think that it is time to move on from basic science and instead to find solutions to actual problems. Both approaches are valid. It is true that there are problems that demand immediate attention; however, few solutions would be found without the help of the wide knowledge foundation generated through basic science.

One example of how basic and applied science can work together to solve practical problems occurred after the discovery of DNA structure led to an understanding of the molecular mechanisms governing DNA replication. Strands of DNA, unique in every human, are found in our cells, where they provide the instructions necessary for life. During DNA replication, DNA makes new copies of itself, shortly before a cell divides. Understanding the mechanisms of DNA replication enabled scientists to develop laboratory techniques that are now used to identify genetic diseases, pinpoint individuals who were at a crime scene, and determine paternity. Without basic science, it is unlikely that applied science would exist.

Another example of the link between basic and applied research is the Human Genome Project, a study in which each human chromosome was analyzed and mapped to determine the precise sequence of DNA subunits and the exact location of each gene. (The gene is the basic unit of heredity; an individual’s complete collection of genes is his or her genome.) Other less complex organisms have also been studied as part of this project in

order to gain a better understanding of human chromosomes. The Human Genome Project ([\[link\]](#)) relied on basic research carried out with simple organisms and, later, with the human genome. An important end goal eventually became using the data for applied research, seeking cures and early diagnoses for genetically related diseases.



The Human Genome Project was a 13-year collaborative effort among researchers working in several different fields of science. The project, which sequenced the entire human genome, was completed in 2003. (credit: the U.S. Department of Energy Genome Programs (<http://genomics.energy.gov>))

While research efforts in both basic science and applied science are usually carefully planned, it is important to note that some discoveries are made by **serendipity**, that is, by means of a fortunate accident or a lucky surprise. Penicillin was discovered when biologist Alexander Fleming accidentally left a petri dish of *Staphylococcus* bacteria open. An unwanted mold grew on the dish, killing the bacteria. The mold turned out to be *Penicillium*, and a new antibiotic was discovered. Even in the highly organized world of science, luck—when combined with an observant, curious mind—can lead to unexpected breakthroughs.

## Reporting Scientific Work

Whether scientific research is basic science or applied science, scientists must share their findings in order for other researchers to expand and build upon their discoveries. Collaboration with other scientists—when planning, conducting, and analyzing results—are all important for scientific research. For this reason, important aspects of a scientist's work are communicating with peers and disseminating results to peers. Scientists can share results by presenting them at a scientific meeting or conference, but this approach can reach only the select few who are present. Instead, most scientists present their results in peer-reviewed manuscripts that are published in scientific journals. **Peer-reviewed manuscripts** are scientific papers that are reviewed by a scientist's colleagues, or peers. These colleagues are qualified individuals, often experts in the same research area, who judge whether or not the scientist's work is suitable for publication. The process of peer review helps to ensure that the research described in a scientific paper or grant proposal is original, significant, logical, and thorough. Grant proposals, which are requests for research funding, are also subject to peer review. Scientists publish their work so other scientists can reproduce their experiments under similar or different conditions to expand on the findings. The experimental results must be consistent with the findings of other scientists.

A scientific paper is very different from creative writing. Although creativity is required to design experiments, there are fixed guidelines when it comes to presenting scientific results. First, scientific writing must be

brief, concise, and accurate. A scientific paper needs to be succinct but detailed enough to allow peers to reproduce the experiments.

The scientific paper consists of several specific sections—introduction, materials and methods, results, and discussion. This structure is sometimes called the “IMRaD” format. There are usually acknowledgment and reference sections as well as an **abstract** (a concise summary) at the beginning of the paper. There might be additional sections depending on the type of paper and the journal where it will be published; for example, some review papers require an outline.

The **introduction** starts with brief, but broad, background information about what is known in the field. A good introduction also gives the rationale of the work; it justifies the work carried out and also briefly mentions the end of the paper, where the hypothesis or research question driving the research will be presented. The introduction refers to the published scientific work of others and therefore requires citations following the style of the journal. Using the work or ideas of others without proper citation is considered **plagiarism**.

The **materials and methods** section includes a complete and accurate description of the substances used, and the method and techniques used by the researchers to gather data. The description should be thorough enough to allow another researcher to repeat the experiment and obtain similar results, but it does not have to be verbose. This section will also include information on how measurements were made and what types of calculations and statistical analyses were used to examine raw data. Although the materials and methods section gives an accurate description of the experiments, it does not discuss them.

Some journals require a results section followed by a discussion section, but it is more common to combine both. If the journal does not allow the combination of both sections, the **results** section simply narrates the findings without any further interpretation. The results are presented by means of tables or graphs, but no duplicate information should be presented. In the **discussion** section, the researcher will interpret the results, describe how variables may be related, and attempt to explain the observations. It is indispensable to conduct an extensive literature search to

put the results in the context of previously published scientific research. Therefore, proper citations are included in this section as well.

Finally, the **conclusion** section summarizes the importance of the experimental findings. While the scientific paper almost certainly answered one or more scientific questions that were stated, any good research should lead to more questions. Therefore, a well-done scientific paper leaves doors open for the researcher and others to continue and expand on the findings.

**Review articles** do not follow the IMRAD format because they do not present original scientific findings, or primary literature; instead, they summarize and comment on findings that were published as primary literature and typically include extensive reference sections.

## Section Summary

Biology is the science that studies living organisms and their interactions with one another and their environments. Science attempts to describe and understand the nature of the universe in whole or in part by rational means. Science has many fields; those fields related to the physical world and its phenomena are considered natural sciences.

Science can be basic or applied. The main goal of basic science is to expand knowledge without any expectation of short-term practical application of that knowledge. The primary goal of applied research, however, is to solve practical problems.

Two types of logical reasoning are used in science. Inductive reasoning uses particular results to produce general scientific principles. Deductive reasoning is a form of logical thinking that predicts results by applying general principles. The common thread throughout scientific research is the use of the scientific method, a step-based process that consists of making observations, defining a problem, posing hypotheses, testing these hypotheses, and drawing one or more conclusions. The testing uses proper controls. Scientists present their results in peer-reviewed scientific papers published in scientific journals. A scientific research paper consists of several well-defined sections: introduction, materials and methods, results,

and, finally, a concluding discussion. Review papers summarize the research done in a particular field over a period of time.

## Art Connections

### Exercise:

#### Problem:

[\[link\]](#) In the example below, the scientific method is used to solve an everyday problem. Order the scientific method steps (numbered items) with the process of solving the everyday problem (lettered items). Based on the results of the experiment, is the hypothesis correct? If it is incorrect, propose some alternative hypotheses.

1. Observation
  2. Question
  3. Hypothesis (answer)
  4. Prediction
  5. Experiment
  6. Result
- 
- a. There is something wrong with the electrical outlet.
  - b. If something is wrong with the outlet, my coffeemaker also won't work when plugged into it.
  - c. My toaster doesn't toast my bread.
  - d. I plug my coffee maker into the outlet.
  - e. My coffeemaker works.
  - f. Why doesn't my toaster work?

---

#### Solution:

[\[link\]](#) 1: C; 2: F; 3: A; 4: B; 5: D; 6: E. The original hypothesis is incorrect, as the coffeemaker works when plugged into the outlet. Alternative hypotheses include that the toaster might be broken or that the toaster wasn't turned on.

### Exercise:



**Problem:**

[\[link\]](#) Decide if each of the following is an example of inductive or deductive reasoning.

1. All flying birds and insects have wings. Birds and insects flap their wings as they move through the air. Therefore, wings enable flight.
2. Insects generally survive mild winters better than harsh ones. Therefore, insect pests will become more problematic if global temperatures increase.
3. Chromosomes, the carriers of DNA, separate into daughter cells during cell division. Therefore, DNA is the genetic material.
4. Animals as diverse as humans, insects, and wolves all exhibit social behavior. Therefore, social behavior must have an evolutionary advantage.

---

**Solution:**

[\[link\]](#) 1: inductive; 2: deductive; 3: deductive; 4: inductive.

**Review Questions****Exercise:**

**Problem:** The first forms of life on Earth were \_\_\_\_\_.

- a. plants
- b. microorganisms
- c. birds
- d. dinosaurs

---

**Solution:**

B

**Exercise:**

**Problem:**

A suggested and testable explanation for an event is called a \_\_\_\_\_.

- a. hypothesis
- b. variable
- c. theory
- d. control

---

**Solution:**

A

**Exercise:**

**Problem:**

Which of the following sciences is not considered a natural science?

- a. biology
- b. astronomy
- c. physics
- d. computer science

---

**Solution:**

D

**Exercise:**

**Problem:**

The type of logical thinking that uses related observations to arrive at a general conclusion is called \_\_\_\_\_.

- a. deductive reasoning
  - b. the scientific method
  - c. hypothesis-based science
  - d. inductive reasoning
- 

**Solution:**

D

**Exercise:**

**Problem:**

The process of \_\_\_\_\_ helps to ensure that a scientist's research is original, significant, logical, and thorough.

- a. publication
  - b. public speaking
  - c. peer review
  - d. the scientific method
- 

**Solution:**

C

**Exercise:**

**Problem:**

A person notices that her houseplants that are regularly exposed to music seem to grow more quickly than those in rooms with no music. As a result, she determines that plants grow better when exposed to music. This example most closely resembles which type of reasoning?

- a. inductive reasoning
- b. deductive reasoning
- c. neither, because no hypothesis was made
- d. both inductive and deductive reasoning

---

**Solution:**

A

**Free Response****Exercise:****Problem:**

Although the scientific method is used by most of the sciences, it can also be applied to everyday situations. Think about a problem that you may have at home, at school, or with your car, and apply the scientific method to solve it.

---

**Solution:**

Answers will vary, but should apply the steps of the scientific method. One possibility could be a car which doesn't start. The hypothesis could be that the car doesn't start because the battery is dead. The experiment would be to change the battery or to charge the battery and then check whether the car starts or not. If it starts, the problem was due to the battery, and the hypothesis is accepted.

**Exercise:****Problem:**

Give an example of how applied science has had a direct effect on your daily life.

---

**Solution:**

Answers will vary. One example of how applied science has had a direct effect on daily life is the presence of vaccines. Vaccines to prevent diseases such as polio, measles, tetanus, and even influenza affect daily life by contributing to individual and societal health.

**Exercise:****Problem:**

Name two topics that are likely to be studied by biologists, and two areas of scientific study that would fall outside the realm of biology.

---

**Solution:**

Answers will vary. Topics that fall inside the area of biological study include how diseases affect human bodies, how pollution impacts a species' habitat, and how plants respond to their environments. Topics that fall outside of biology (the "study of life") include how metamorphic rock is formed and how planetary orbits function.

**Exercise:****Problem:**

Thinking about the topic of cancer, write a basic science question and an applied science question that a researcher interested in this topic might ask

---

**Solution:**

Answers will vary. Basic science: What evolutionary purpose might cancer serve? Applied science: What strategies might be found to prevent cancer from reproducing at the cellular level?

**Glossary**

abstract

opening section of a scientific paper that summarizes the research and conclusions

applied science

form of science that aims to solve real-world problems

basic science

science that seeks to expand knowledge and understanding regardless of the short-term application of that knowledge

biology

the study of living organisms and their interactions with one another and their environments

conclusion

section of a scientific paper that summarizes the importance of the experimental findings

control

part of an experiment that does not change during the experiment

deductive reasoning

form of logical thinking that uses a general inclusive statement to forecast specific results

descriptive science

(also, discovery science) form of science that aims to observe, explore, and investigate

discussion

section of a scientific paper in which the author interprets experimental results, describes how variables may be related, and attempts to explain the phenomenon in question

falsifiable

able to be disproven by experimental results

hypothesis

suggested explanation for an observation, which can be tested

hypothesis-based science

form of science that begins with a specific question and potential testable answers

inductive reasoning

form of logical thinking that uses related observations to arrive at a general conclusion

introduction

opening section of a scientific paper, which provides background information about what was known in the field prior to the research reported in the paper

life science

field of science, such as biology, that studies living things

materials and methods

section of a scientific paper that includes a complete description of the substances, methods, and techniques used by the researchers to gather data

natural science

field of science that is related to the physical world and its phenomena and processes

peer-reviewed manuscript

scientific paper that is reviewed by a scientist's colleagues who are experts in the field of study

physical science

field of science, such as geology, astronomy, physics, and chemistry, that studies nonliving matter

plagiarism

using other people's work or ideas without proper citation, creating the false impression that those are the author's original ideas

results

section of a scientific paper in which the author narrates the experimental findings and presents relevant figures, pictures, diagrams, graphs, and tables, without any further interpretation

review article

paper that summarizes and comments on findings that were published as primary literature

science

knowledge that covers general truths or the operation of general laws, especially when acquired and tested by the scientific method

scientific method

method of research with defined steps that include observation, formulation of a hypothesis, testing, and confirming or falsifying the hypothesis

serendipity

fortunate accident or a lucky surprise

theory

tested and confirmed explanation for observations or phenomena

variable

part of an experiment that the experimenter can vary or change



## Themes and Concepts of Biology

By the end of this section, you will be able to:

- Identify and describe the properties of life
- Describe the levels of organization among living things
- Recognize and interpret a phylogenetic tree
- List examples of different sub disciplines in biology

Biology is the science that studies life, but what exactly is life? This may sound like a silly question with an obvious response, but it is not always easy to define life. For example, a branch of biology called virology studies viruses, which exhibit some of the characteristics of living entities but lack others. It turns out that although viruses can attack living organisms, cause diseases, and even reproduce, they do not meet the criteria that biologists use to define life. Consequently, virologists are not biologists, strictly speaking. Similarly, some biologists study the early molecular evolution that gave rise to life; since the events that preceded life are not biological events, these scientists are also excluded from biology in the strict sense of the term.

From its earliest beginnings, biology has wrestled with three questions: What are the shared properties that make something “alive”? And once we know something is alive, how do we find meaningful levels of organization in its structure? And, finally, when faced with the remarkable diversity of life, how do we organize the different kinds of organisms so that we can better understand them? As new organisms are discovered every day, biologists continue to seek answers to these and other questions.

## Properties of Life

All living organisms share several key characteristics or functions: order, sensitivity or response to the environment, reproduction, adaptation, growth and development, regulation, homeostasis, energy processing, and evolution. When viewed together, these nine characteristics serve to define life.

## Order



A toad represents a highly organized structure consisting of cells, tissues, organs, and organ systems. (credit: “Ivengo”/Wikimedia Commons)

Organisms are highly organized, coordinated structures that consist of one or more cells. Even very simple, single-celled organisms are remarkably complex: inside each cell, atoms make up molecules; these in turn make up cell organelles and other cellular inclusions. In multicellular organisms ([link](#)), similar cells form tissues. Tissues, in turn, collaborate to create organs (body structures with a distinct function). Organs work together to form organ systems.

## Sensitivity or Response to Stimuli



The leaves of this sensitive plant (*Mimosa pudica*) will instantly droop and fold when touched. After a few minutes, the plant returns to normal. (credit: Alex Lomas)

Organisms respond to diverse stimuli. For example, plants can bend toward a source of light, climb on fences and walls, or respond to touch ([\[link\]](#)). Even tiny bacteria can move toward or away from chemicals (a process called *chemotaxis*) or light (*phototaxis*). Movement toward a stimulus is considered a positive response, while movement away from a stimulus is considered a negative response.

**Note:**

Link to Learning



Watch [this video](#) to see how plants respond to a stimulus—from opening to light, to wrapping a tendril around a branch, to capturing prey.

## **Reproduction**

Single-celled organisms reproduce by first duplicating their DNA, and then dividing it equally as the cell prepares to divide to form two new cells. Multicellular organisms often produce specialized reproductive germline cells that will form new individuals. When reproduction occurs, genes containing DNA are passed along to an organism's offspring. These genes ensure that the offspring will belong to the same species and will have similar characteristics, such as size and shape.

## **Growth and Development**

Organisms grow and develop following specific instructions coded for by their genes. These genes provide instructions that will direct cellular growth and development, ensuring that a species' young ([link](#)) will grow up to exhibit many of the same characteristics as its parents.



Although no two look alike, these kittens have inherited genes from both parents and share many of the same characteristics. (credit: Rocky Mountain Feline Rescue)

## **Regulation**

Even the smallest organisms are complex and require multiple regulatory mechanisms to coordinate internal functions, respond to stimuli, and cope with environmental stresses. Two examples of internal functions regulated in an organism are nutrient transport and blood flow. Organs (groups of tissues working together) perform specific functions, such as carrying oxygen throughout the body, removing wastes, delivering nutrients to every cell, and cooling the body.

## **Homeostasis**



Polar bears (*Ursus maritimus*) and other mammals living in ice-covered regions maintain their body temperature by generating heat and reducing heat loss through thick fur and a dense layer of fat under their skin. (credit: “longhorndave”/Flickr)

In order to function properly, cells need to have appropriate conditions such as proper temperature, pH, and appropriate concentration of diverse chemicals. These conditions may, however, change from one moment to the next. Organisms are able to maintain internal conditions within a narrow range almost constantly, despite environmental changes, through **homeostasis** (literally, “steady state”)—the ability of an organism to maintain constant internal conditions. For example, an organism needs to regulate body temperature through a process known as thermoregulation. Organisms that live in cold climates, such as the polar bear ([\[link\]](#)), have body structures that help them withstand low temperatures and conserve body heat. Structures that aid in this type of insulation include fur, feathers, blubber, and fat. In hot climates, organisms have methods (such as perspiration in humans or panting in dogs) that help them to shed excess body heat.

## Energy Processing



The California condor (*Gymnogyps californianus*) uses chemical energy derived from food to power flight. California condors are an endangered species; this bird has a wing tag that helps biologists identify the individual. (credit: Pacific Southwest Region U.S. Fish and Wildlife Service)

All organisms use a source of energy for their metabolic activities. Some organisms capture energy from the sun and convert it into chemical energy

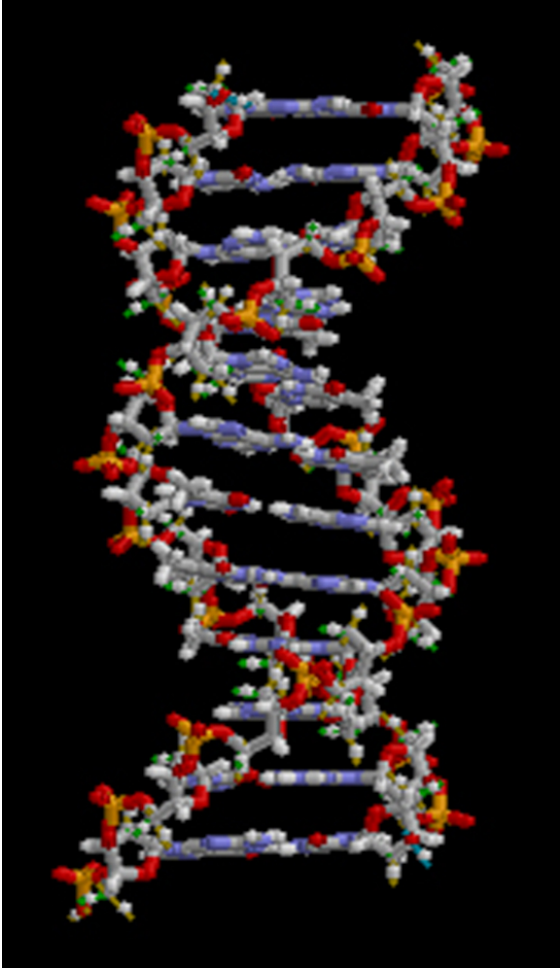


in food; others use chemical energy in molecules they take in as food ([link](#)).

## Levels of Organization of Living Things

Living things are highly organized and structured, following a hierarchy that can be examined on a scale from small to large. The **atom** is the smallest and most fundamental unit of matter. It consists of a nucleus surrounded by electrons. Atoms form molecules. A **molecule** is a chemical structure consisting of at least two atoms held together by one or more chemical bonds. Many molecules that are biologically important are **macromolecules**, large molecules that are typically formed by polymerization (a polymer is a large molecule that is made by combining smaller units called monomers, which are simpler than macromolecules). An example of a macromolecule is deoxyribonucleic acid (DNA) ([link](#)), which contains the instructions for the structure and functioning of all living organisms.





All molecules, including this  
DNA molecule, are  
composed of atoms. (credit:  
“brian0918”/Wikimedia  
Commons)

**Note:**  
Link to Learning



Watch [this video](#) that animates the three-dimensional structure of the DNA molecule shown in [\[link\]](#).

Some cells contain aggregates of macromolecules surrounded by membranes; these are called **organelles**. Organelles are small structures that exist within cells. Examples of organelles include mitochondria and chloroplasts, which carry out indispensable functions: mitochondria produce energy to power the cell, while chloroplasts enable green plants to utilize the energy in sunlight to make sugars. All living things are made of cells; the **cell** itself is the smallest fundamental unit of structure and function in living organisms. (This requirement is why viruses are not considered living: they are not made of cells. To make new viruses, they have to invade and hijack the reproductive mechanism of a living cell; only then can they obtain the materials they need to reproduce.) Some organisms consist of a single cell and others are multicellular. Cells are classified as prokaryotic or eukaryotic. **Prokaryotes** are single-celled or colonial organisms that do not have membrane-bound nuclei; in contrast, the cells of **eukaryotes** do have membrane-bound organelles and a membrane-bound nucleus.

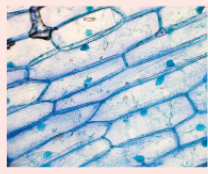
In larger organisms, cells combine to make **tissues**, which are groups of similar cells carrying out similar or related functions. **Organs** are collections of tissues grouped together performing a common function. Organs are present not only in animals but also in plants. An **organ system** is a higher level of organization that consists of functionally related organs. Mammals have many organ systems. For instance, the circulatory system transports blood through the body and to and from the lungs; it includes organs such as the heart and blood vessels. **Organisms** are individual living entities. For example, each tree in a forest is an organism. Single-celled

prokaryotes and single-celled eukaryotes are also considered organisms and are typically referred to as microorganisms.

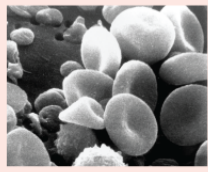
All the individuals of a species living within a specific area are collectively called a **population**. For example, a forest may include many pine trees. All of these pine trees represent the population of pine trees in this forest. Different populations may live in the same specific area. For example, the forest with the pine trees includes populations of flowering plants and also insects and microbial populations. A **community** is the sum of populations inhabiting a particular area. For instance, all of the trees, flowers, insects, and other populations in a forest form the forest's community. The forest itself is an ecosystem. An **ecosystem** consists of all the living things in a particular area together with the abiotic, non-living parts of that environment such as nitrogen in the soil or rain water. At the highest level of organization ([\[link\]](#)), the **biosphere** is the collection of all ecosystems, and it represents the zones of life on earth. It includes land, water, and even the atmosphere to a certain extent.

**Note:**

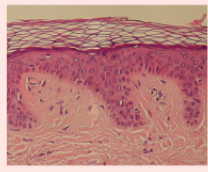
Art Connection



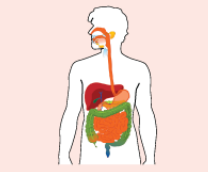
**Organelles:** The nucleus, dyed blue in these onion cells, is an example of an organelle.



**Cells:** Human blood cells.



**Tissues:** Human skin tissue.



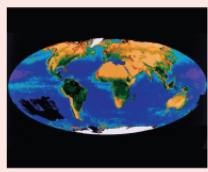
**Organs and Organ Systems:** Organs, such as the stomach and intestine, make up the human digestive system.



**Organisms, Populations, and Communities:** In a forest, each pine tree is an organism. Together, all the pine trees make up a population. All the plant and animal species in the forest comprise a community.



**Ecosystems:** This coastal ecosystem in the southeastern United States includes living organisms and the environment in which they live.



**The Biosphere:** Encompasses all the ecosystems on Earth.

The biological levels of organization of living things are shown. From a single organelle to the

entire biosphere, living organisms are parts of a highly structured hierarchy. (credit “organelles”: modification of work by Umberto Salvagnin; credit “cells”: modification of work by Bruce Wetzel, Harry Schaefer/ National Cancer Institute; credit “tissues”: modification of work by Kilbad; Fama Clamosa; Mikael Häggström; credit “organs”: modification of work by Mariana Ruiz Villareal; credit “organisms”: modification of work by "Crystal"/Flickr; credit “ecosystems”: modification of work by US Fish and Wildlife Service Headquarters; credit “biosphere”: modification of work by NASA)

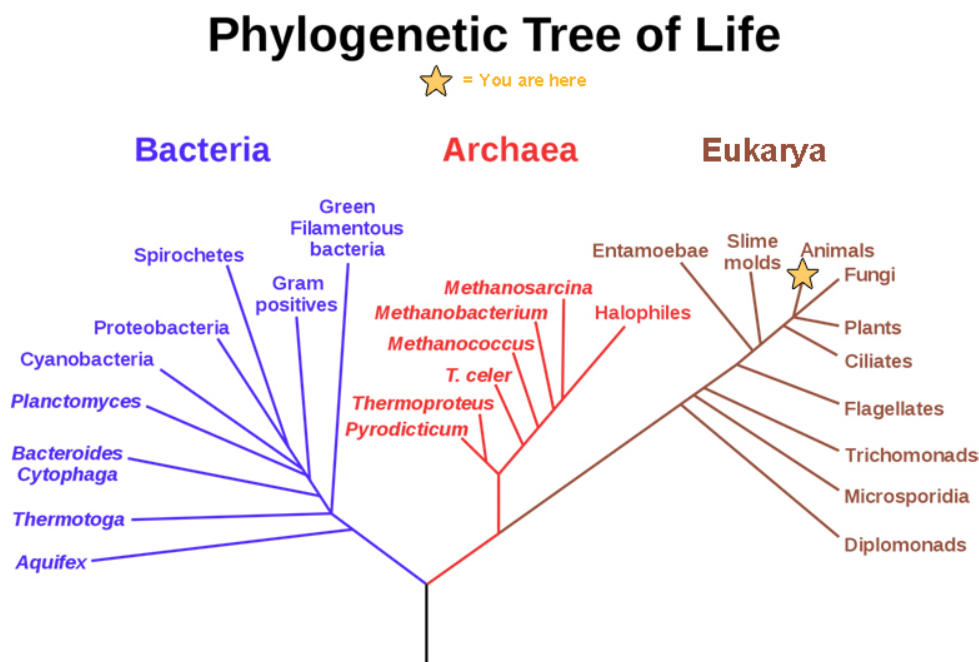
Which of the following statements is false?

- a. Tissues exist within organs which exist within organ systems.
- b. Communities exist within populations which exist within ecosystems.
- c. Organelles exist within cells which exist within tissues.
- d. Communities exist within ecosystems which exist in the biosphere.

## The Diversity of Life

The fact that biology, as a science, has such a broad scope has to do with the tremendous diversity of life on earth. The source of this diversity is **evolution**, the process of gradual change during which new species arise from older species. Evolutionary biologists study the evolution of living things in everything from the microscopic world to ecosystems.

The evolution of various life forms on Earth can be summarized in a phylogenetic tree ([link](#)). A **phylogenetic tree** is a diagram showing the evolutionary relationships among biological species based on similarities and differences in genetic or physical traits or both. A phylogenetic tree is composed of nodes and branches. The internal nodes represent ancestors and are points in evolution when, based on scientific evidence, an ancestor is thought to have diverged to form two new species. The length of each branch is proportional to the time elapsed since the split.



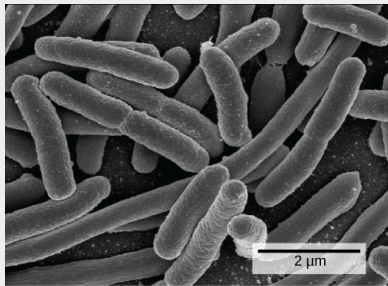
This phylogenetic tree was constructed by microbiologist Carl Woese using data obtained from sequencing ribosomal RNA genes. The tree shows the separation of living organisms into three domains: Bacteria, Archaea, and Eukarya. Bacteria and Archaea are prokaryotes, single-celled organisms lacking intracellular organelles. (credit: Eric Gaba; NASA Astrobiology Institute)

**Note:****Evolution Connection****Carl Woese and the Phylogenetic Tree**

In the past, biologists grouped living organisms into five kingdoms: animals, plants, fungi, protists, and bacteria. The organizational scheme was based mainly on physical features, as opposed to physiology, biochemistry, or molecular biology, all of which are used by modern systematics. The pioneering work of American microbiologist Carl Woese in the early 1970s has shown, however, that life on Earth has evolved along three lineages, now called domains—Bacteria, Archaea, and Eukarya. The first two are prokaryotic cells with microbes that lack membrane-enclosed nuclei and organelles. The third domain contains the eukaryotes and includes unicellular microorganisms together with the four original kingdoms (excluding bacteria). Woese defined Archaea as a new domain, and this resulted in a new taxonomic tree ([\[link\]](#)). Many organisms belonging to the Archaea domain live under extreme conditions and are called extremophiles. To construct his tree, Woese used genetic relationships rather than similarities based on morphology (shape). Woese's tree was constructed from comparative sequencing of the genes that are universally distributed, present in every organism, and conserved (meaning that these genes have remained essentially unchanged throughout evolution). Woese's approach was revolutionary because comparisons of physical features are insufficient to differentiate between the prokaryotes that appear fairly similar in spite of their tremendous biochemical diversity and genetic variability ([\[link\]](#)). The comparison of homologous DNA and

RNA sequences provided Woese with a sensitive device that revealed the extensive variability of prokaryotes, and which justified the separation of the prokaryotes into two domains: bacteria and archaea.

These images represent different domains. The (a) bacteria in this micrograph belong to Domain Bacteria, while the (b) extremophiles (not visible) living in this hot vent belong to Domain Archaea. Both the (c) sunflower and (d) lion are part of Domain Eukarya. (credit a: modification of work by Drew March; credit b: modification of work by Steve Jurvetson; credit c: modification of work by Michael Arrighi; credit d: modification of work by Leszek Leszcynski)



(a)



(b)



(c)



(d)

## Branches of Biological Study

The scope of biology is broad and therefore contains many branches and subdisciplines. Biologists may pursue one of those subdisciplines and work



in a more focused field. For instance, **molecular biology** and **biochemistry** study biological processes at the molecular and chemical level, including interactions among molecules such as DNA, RNA, and proteins, as well as the way they are regulated. **Microbiology**, the study of microorganisms, is the study of the structure and function of single-celled organisms. It is quite a broad branch itself, and depending on the subject of study, there are also microbial physiologists, ecologists, and geneticists, among others.

**Note:****Career Connection****Forensic Scientist**

Forensic science is the application of science to answer questions related to the law. Biologists as well as chemists and biochemists can be forensic scientists. Forensic scientists provide scientific evidence for use in courts, and their job involves examining trace materials associated with crimes. Interest in forensic science has increased in the last few years, possibly because of popular television shows that feature forensic scientists on the job. Also, the development of molecular techniques and the establishment of DNA databases have expanded the types of work that forensic scientists can do. Their job activities are primarily related to crimes against people such as murder, rape, and assault. Their work involves analyzing samples such as hair, blood, and other body fluids and also processing DNA ([\[link\]](#)) found in many different environments and materials. Forensic scientists also analyze other biological evidence left at crime scenes, such as insect larvae or pollen grains. Students who want to pursue careers in forensic science will most likely be required to take chemistry and biology courses as well as some intensive math courses.



This forensic scientist works in a DNA extraction room at the U.S. Army Criminal Investigation Laboratory at Fort Gillem, GA. (credit: United States Army CID Command Public Affairs)

Another field of biological study, **neurobiology**, studies the biology of the nervous system, and although it is considered a branch of biology, it is also recognized as an interdisciplinary field of study known as neuroscience. Because of its interdisciplinary nature, this subdiscipline studies different functions of the nervous system using molecular, cellular, developmental, medical, and computational approaches.



Researchers work on excavating dinosaur fossils at a site in Castellón, Spain. (credit: Mario Modesto)

**Paleontology**, another branch of biology, uses fossils to study life's history ([link](#)). **Zoology** and **botany** are the study of animals and plants, respectively. Biologists can also specialize as biotechnologists, ecologists, or physiologists, to name just a few areas. This is just a small sample of the many fields that biologists can pursue.

Biology is the culmination of the achievements of the natural sciences from their inception to today. Excitingly, it is the cradle of emerging sciences, such as the biology of brain activity, genetic engineering of custom organisms, and the biology of evolution that uses the laboratory tools of molecular biology to retrace the earliest stages of life on earth. A scan of news headlines—whether reporting on immunizations, a newly discovered species, sports doping, or a genetically-modified food—demonstrates the way biology is active in and important to our everyday world.

## Section Summary

Biology is the science of life. All living organisms share several key properties such as order, sensitivity or response to stimuli, reproduction, growth and development, regulation, homeostasis, and energy processing. Living things are highly organized parts of a hierarchy that includes atoms, molecules, organelles, cells, tissues, organs, and organ systems. Organisms, in turn, are grouped as populations, communities, ecosystems, and the biosphere. The great diversity of life today evolved from less-diverse ancestral organisms over billions of years. A diagram called a phylogenetic tree can be used to show evolutionary relationships among organisms.

Biology is very broad and includes many branches and subdisciplines. Examples include molecular biology, microbiology, neurobiology, zoology, and botany, among others.

## Art Connections

### Exercise:

**Problem:** [\[link\]](#) Which of the following statements is false?

- a. Tissues exist within organs which exist within organ systems.
- b. Communities exist within populations which exist within ecosystems.
- c. Organelles exist within cells which exist within tissues.
- d. Communities exist within ecosystems which exist in the biosphere.

---

### Solution:

[\[link\]](#) Communities exist within populations which exist within ecosystems.

## Review Questions

### Exercise:

**Problem:**

The smallest unit of biological structure that meets the functional requirements of “living” is the \_\_\_\_\_.

- a. organ
- b. organelle
- c. cell
- d. macromolecule

---

**Solution:**

C

**Exercise:**

**Problem:** Viruses are not considered living because they \_\_\_\_\_.

- a. are not made of cells
- b. lack cell nuclei
- c. do not contain DNA or RNA
- d. cannot reproduce

---

**Solution:**

A

**Exercise:**

**Problem:**

The presence of a membrane-enclosed nucleus is a characteristic of \_\_\_\_\_.

- a. prokaryotic cells
- b. eukaryotic cells
- c. living organisms

d. bacteria

---

**Solution:**

B

**Exercise:**

**Problem:**

A group of individuals of the same species living in the same area is called a(n) \_\_\_\_\_.

- a. family
- b. community
- c. population
- d. ecosystem

---

**Solution:**

C

**Exercise:**

**Problem:**

Which of the following sequences represents the hierarchy of biological organization from the most inclusive to the least complex level?

- a. organelle, tissue, biosphere, ecosystem, population
- b. organ, organism, tissue, organelle, molecule
- c. organism, community, biosphere, molecule, tissue, organ
- d. biosphere, ecosystem, community, population, organism

---

**Solution:**

D

**Exercise:****Problem:**

Where in a phylogenetic tree would you expect to find the organism that had evolved most recently?

- a. at the base
- b. within the branches
- c. at the nodes
- d. at the branch tips

---

**Solution:**

D

**Free Response****Exercise:****Problem:**

Select two items that biologists agree are necessary in order to consider an organism “alive.” For each, give an example of a non-living object that otherwise fits the definition of “alive,”

---

**Solution:**

Answers will vary. Layers of sedimentary rock have order but are not alive. Technology is capable of regulation but is not, of itself, alive.

**Exercise:**

**Problem:**

Consider the levels of organization of the biological world, and place each of these items in order from smallest level of organization to most encompassing: skin cell, elephant, water molecule, planet Earth, tropical rainforest, hydrogen atom, wolf pack, liver.

---

**Solution:**

Smallest level of organization to largest: hydrogen atom, water molecule, skin cell, liver, elephant, wolf pack, tropical rainforest, planet Earth

**Exercise:****Problem:**

You go for a long walk on a hot day. Give an example of a way in which homeostasis keeps your body healthy.

---

**Solution:**

During your walk, you may begin to perspire, which cools your body and helps your body to maintain a constant internal temperature. You might also become thirsty and pause long enough for a cool drink, which will help to restore the water lost during perspiration.

**Exercise:****Problem:**

Using examples, explain how biology can be studied from a microscopic approach to a global approach.

---

**Solution:**

Researchers can approach biology from the smallest to the largest, and everything in between. For instance, an ecologist may study a population of individuals, the population's community, the community's ecosystem, and the ecosystem's part in the biosphere.



When studying an individual organism, a biologist could examine the cell and its organelles, the tissues that the cells make up, the organs and their respective organ systems, and the sum total—the organism itself.

## **Glossary**

atom

smallest and most fundamental unit of matter

biochemistry

study of the chemistry of biological organisms

biosphere

collection of all the ecosystems on Earth

botany

study of plants

cell

smallest fundamental unit of structure and function in living things

community

set of populations inhabiting a particular area

ecosystem

all the living things in a particular area together with the abiotic, nonliving parts of that environment

eukaryote

organism with cells that have nuclei and membrane-bound organelles

evolution

process of gradual change during which new species arise from older species and some species become extinct

homeostasis

ability of an organism to maintain constant internal conditions

macromolecule

large molecule, typically formed by the joining of smaller molecules

microbiology

study of the structure and function of microorganisms

molecule

chemical structure consisting of at least two atoms held together by one or more chemical bonds

molecular biology

study of biological processes and their regulation at the molecular level, including interactions among molecules such as DNA, RNA, and proteins

neurobiology

study of the biology of the nervous system

organ

collection of related tissues grouped together performing a common function

organ system

level of organization that consists of functionally related interacting organs

organelle

small structures that exist within cells and carry out cellular functions

organism

individual living entity

paleontology

study of life's history by means of fossils

phylogenetic tree

diagram showing the evolutionary relationships among various biological species based on similarities and differences in genetic or physical traits or both; in essence, a hypothesis concerning evolutionary connections

population

all of the individuals of a species living within a specific area

prokaryote

single-celled organism that lacks organelles and does not have nuclei surrounded by a nuclear membrane

tissue

group of similar cells carrying out related functions

zoology

study of animals

## Introduction

class="introduction"

An arctic fox  
is a complex  
animal, well  
adapted to  
its  
environment  
. It changes  
coat color  
with the  
seasons, and  
has longer  
fur in winter  
to trap heat.  
(credit:  
modification  
of work by  
Keith  
Morehouse,  
USFWS)



The arctic fox is an example of a complex animal that has adapted to its environment and illustrates the relationships between an animal's form and function. The structures of animals consist of primary tissues that make up more complex organs and organ systems. Homeostasis allows an animal to maintain a balance between its internal and external environments.

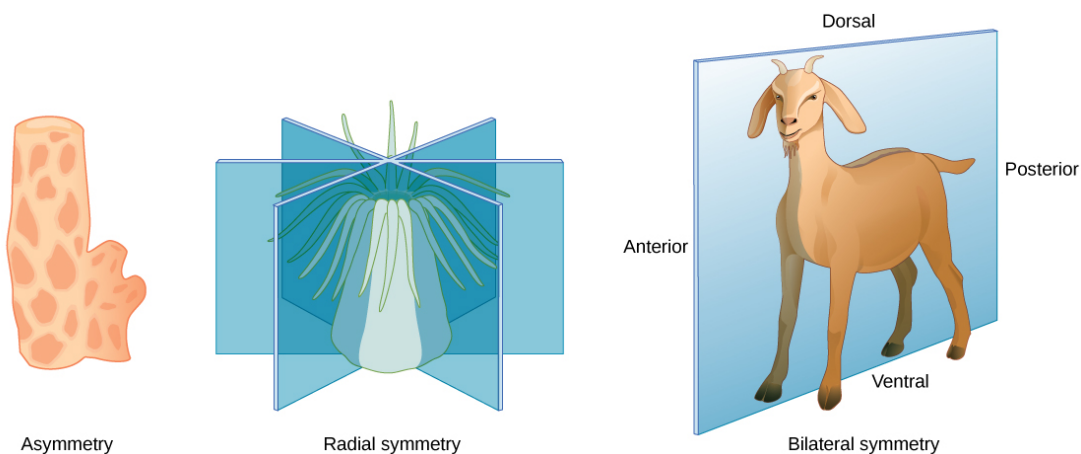
## Animal Form and Function

By the end of this section, you will be able to:

- Describe the various types of body plans that occur in animals
- Describe limits on animal size and shape
- Relate bioenergetics to body size, levels of activity, and the environment

Animals vary in form and function. From a sponge to a worm to a goat, an organism has a distinct body plan that limits its size and shape. Animals' bodies are also designed to interact with their environments, whether in the deep sea, a rainforest canopy, or the desert. Therefore, a large amount of information about the structure of an organism's body (anatomy) and the function of its cells, tissues and organs (physiology) can be learned by studying that organism's environment.

## Body Plans



Animals exhibit different types of body symmetry. The sponge is asymmetrical, the sea anemone has radial symmetry, and the goat has bilateral symmetry.

Animal body plans follow set patterns related to symmetry. They are asymmetrical, radial, or bilateral in form as illustrated in [\[link\]](#).

**Asymmetrical** animals are animals with no pattern or symmetry; an example of an asymmetrical animal is a sponge. Radial symmetry, as illustrated in [\[link\]](#), describes when an animal has an up-and-down orientation: any plane cut along its longitudinal axis through the organism produces equal halves, but not a definite right or left side. This plan is found mostly in aquatic animals, especially organisms that attach themselves to a base, like a rock or a boat, and extract their food from the surrounding water as it flows around the organism. Bilateral symmetry is illustrated in the same figure by a goat. The goat also has an upper and lower component to it, but a plane cut from front to back separates the animal into definite right and left sides. Additional terms used when describing positions in the body are anterior (front), posterior (rear), dorsal (toward the back), and ventral (toward the stomach). Bilateral symmetry is found in both land-based and aquatic animals; it enables a high level of mobility.

## Limits on Animal Size and Shape

Animals with bilateral symmetry that live in water tend to have a **fusiform** shape: this is a tubular shaped body that is tapered at both ends. This shape decreases the drag on the body as it moves through water and allows the animal to swim at high speeds. [\[link\]](#) lists the maximum speed of various animals. Certain types of sharks can swim at fifty kilometers an hour and some dolphins at 32 to 40 kilometers per hour. Land animals frequently travel faster, although the tortoise and snail are significantly slower than cheetahs. Another difference in the adaptations of aquatic and land-dwelling organisms is that aquatic organisms are constrained in shape by the forces of drag in the water since water has higher viscosity than air. On the other hand, land-dwelling organisms are constrained mainly by gravity, and drag is relatively unimportant. For example, most adaptations in birds are for gravity not for drag.

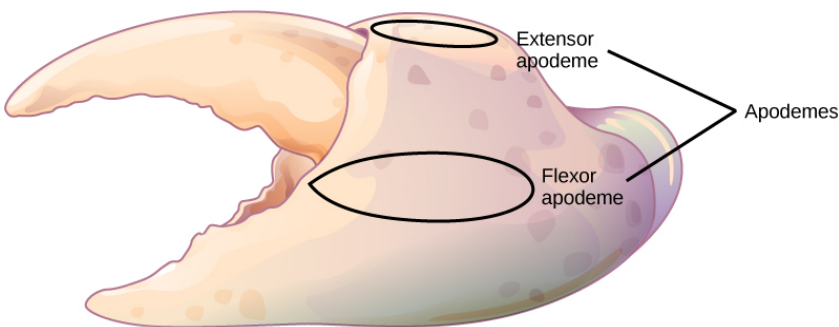
Maximum Speed of Assorted Land Marine Animals		
Animal	Speed (kmh)	Speed (mph)
Cheetah	113	70
Quarter horse	77	48
Fox	68	42
Shortfin mako shark	50	31
Domestic house cat	48	30
Human	45	28
Dolphin	32–40	20–25
Mouse	13	8
Snail	0.05	0.03

Most animals have an exoskeleton, including insects, spiders, scorpions, horseshoe crabs, centipedes, and crustaceans. Scientists estimate that, of insects alone, there are over 30 million species on our planet. The exoskeleton is a hard covering or shell that provides benefits to the animal, such as protection against damage from predators and from water loss (for land animals); it also provides for the attachments of muscles.

As the tough and resistant outer cover of an arthropod, the exoskeleton may be constructed of a tough polymer such as chitin and is often biomineralized with materials such as calcium carbonate. This is fused to the animal's epidermis. Ingrowths of the exoskeleton, called **apodemes**, function as attachment sites for muscles, similar to tendons in more advanced animals ([\[link\]](#)). In order to grow, the animal must first synthesize a new exoskeleton underneath the old one and then shed or molt the original



covering. This limits the animal's ability to grow continually, and may limit the individual's ability to mature if molting does not occur at the proper time. The thickness of the exoskeleton must be increased significantly to accommodate any increase in weight. It is estimated that a doubling of body size increases body weight by a factor of eight. The increasing thickness of the chitin necessary to support this weight limits most animals with an exoskeleton to a relatively small size. The same principles apply to endoskeletons, but they are more efficient because muscles are attached on the outside, making it easier to compensate for increased mass.



Apodemes are ingrowths on arthropod exoskeletons to which muscles attach. The apodemes on this crab leg are located above and below the fulcrum of the claw. Contraction of muscles attached to the apodemes pulls the claw closed.

An animal with an endoskeleton has its size determined by the amount of skeletal system it needs in order to support the other tissues and the amount of muscle it needs for movement. As the body size increases, both bone and muscle mass increase. The speed achievable by the animal is a balance between its overall size and the bone and muscle that provide support and movement.

## **Limiting Effects of Diffusion on Size and Development**

The exchange of nutrients and wastes between a cell and its watery environment occurs through the process of diffusion. All living cells are bathed in liquid, whether they are in a single-celled organism or a multicellular one. Diffusion is effective over a specific distance and limits the size that an individual cell can attain. If a cell is a single-celled microorganism, such as an amoeba, it can satisfy all of its nutrient and waste needs through diffusion. If the cell is too large, then diffusion is ineffective and the center of the cell does not receive adequate nutrients nor is it able to effectively dispel its waste.

An important concept in understanding how efficient diffusion is as a means of transport is the surface to volume ratio. Recall that any three-dimensional object has a surface area and volume; the ratio of these two quantities is the surface-to-volume ratio. Consider a cell shaped like a perfect sphere: it has a surface area of  $4\pi r^2$ , and a volume of  $(4/3)\pi r^3$ . The surface-to-volume ratio of a sphere is  $3/r$ ; as the cell gets bigger, its surface to volume ratio decreases, making diffusion less efficient. The larger the size of the sphere, or animal, the less surface area for diffusion it possesses.

The solution to producing larger organisms is for them to become multicellular. Specialization occurs in complex organisms, allowing cells to become more efficient at doing fewer tasks. For example, circulatory systems bring nutrients and remove waste, while respiratory systems provide oxygen for the cells and remove carbon dioxide from them. Other organ systems have developed further specialization of cells and tissues and efficiently control body functions. Moreover, surface-to-volume ratio applies to other areas of animal development, such as the relationship between muscle mass and cross-sectional surface area in supporting skeletons, and in the relationship between muscle mass and the generation of dissipation of heat.

**Note:**

[Link to Learning](#)



Visit [this interactive site](#) to see an entire animal (a zebrafish embryo) at the cellular and sub-cellular level. Use the zoom and navigation functions for a virtual nanoscopy exploration.



## Animal Bioenergetics

All animals must obtain their energy from food they ingest or absorb. These nutrients are converted to adenosine triphosphate (ATP) for short-term storage and use by all cells. Some animals store energy for slightly longer times as glycogen, and others store energy for much longer times in the form of triglycerides housed in specialized adipose tissues. No energy system is one hundred percent efficient, and an animal's metabolism produces waste energy in the form of heat. If an animal can conserve that heat and maintain a relatively constant body temperature, it is classified as a warm-blooded animal and called an **endotherm**. The insulation used to conserve the body heat comes in the forms of fur, fat, or feathers. The absence of insulation in **ectothermic** animals increases their dependence on the environment for body heat.

The amount of energy expended by an animal over a specific time is called its metabolic rate. The rate is measured variously in joules, calories, or kilocalories (1000 calories). Carbohydrates and proteins contain about 4.5 to 5 kcal/g, and fat contains about 9 kcal/g. Metabolic rate is estimated as the **basal metabolic rate (BMR)** in endothermic animals at rest and as the **standard metabolic rate (SMR)** in ectotherms. Human males have a BMR of 1600 to 1800 kcal/day, and human females have a BMR of 1300 to 1500 kcal/day. Even with insulation, endothermal animals require extensive amounts of energy to maintain a constant body temperature. An ectotherm such as an alligator has an SMR of 60 kcal/day.

## Energy Requirements Related to Body Size

Smaller endothermic animals have a greater surface area for their mass than larger ones ([\[link\]](#)). Therefore, smaller animals lose heat at a faster rate than larger animals and require more energy to maintain a constant internal temperature. This results in a smaller endothermic animal having a higher BMR, per body weight, than a larger endothermic animal.

Species		
Mass	35 g	4,500,000 g
Metabolic rate	890 mm <sup>3</sup> O <sub>2</sub> /g body mass/hr	75 mm <sup>3</sup> O <sub>2</sub> /g body mass/hr

The mouse has a much higher metabolic rate than the elephant. (credit “mouse”: modification of work by Magnus Kjaergaard; credit “elephant”: modification of work by “TheLizardQueen”/Flickr)

## Energy Requirements Related to Levels of Activity

The more active an animal is, the more energy is needed to maintain that activity, and the higher its BMR or SMR. The average daily rate of energy consumption is about two to four times an animal’s BMR or SMR. Humans are more sedentary than most animals and have an average daily rate of only 1.5 times the BMR. The diet of an endothermic animal is determined by its BMR. For example: the type of grasses, leaves, or shrubs that an herbivore eats affects the number of calories that it takes in. The relative caloric content of herbivore foods, in descending order, is tall grasses >

legumes > short grasses > forbs (any broad-leaved plant, not a grass) > subshrubs > annuals/biennials.

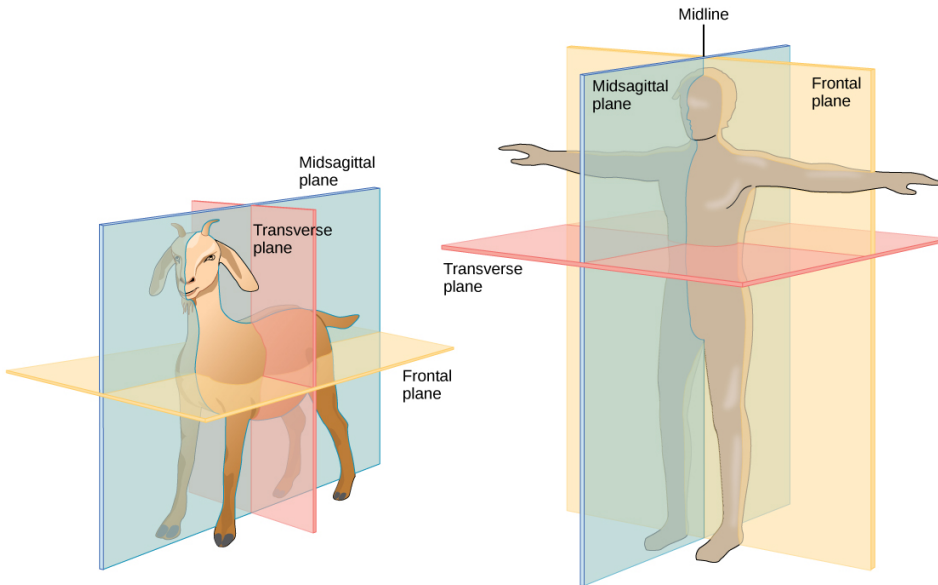
## Energy Requirements Related to Environment

Animals adapt to extremes of temperature or food availability through torpor. **Torpor** is a process that leads to a decrease in activity and metabolism and allows animals to survive adverse conditions. Torpor can be used by animals for long periods, such as entering a state of **hibernation** during the winter months, in which case it enables them to maintain a reduced body temperature. During hibernation, ground squirrels can achieve an abdominal temperature of 0° C (32° F), while a bear's internal temperature is maintained higher at about 37° C (99° F).

If torpor occurs during the summer months with high temperatures and little water, it is called **estivation**. Some desert animals use this to survive the harshest months of the year. Torpor can occur on a daily basis; this is seen in bats and hummingbirds. While endothermy is limited in smaller animals by surface to volume ratio, some organisms can be smaller and still be endotherms because they employ daily torpor during the part of the day that is coldest. This allows them to conserve energy during the colder parts of the day, when they consume more energy to maintain their body temperature.

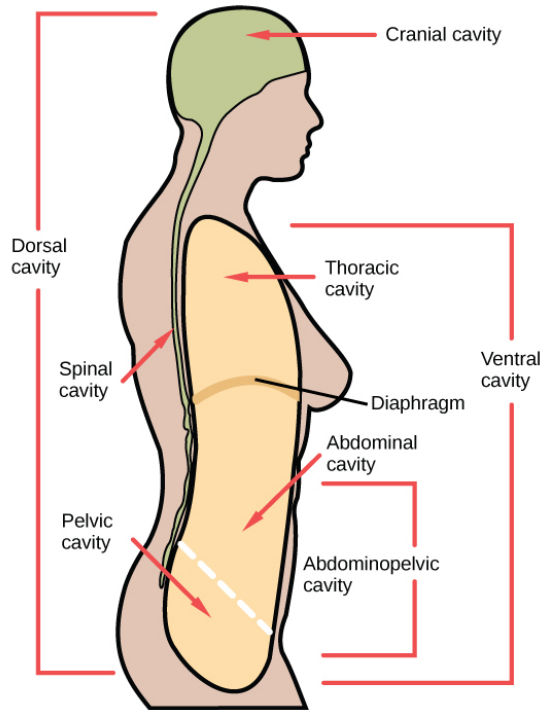
## Animal Body Planes and Cavities

A standing vertebrate animal can be divided by several planes. A **sagittal plane** divides the body into right and left portions. A **midsagittal plane** divides the body exactly in the middle, making two equal right and left halves. A **frontal plane** (also called a coronal plane) separates the front from the back. A **transverse plane** (or, horizontal plane) divides the animal into upper and lower portions. This is sometimes called a cross section, and, if the transverse cut is at an angle, it is called an oblique plane. [\[link\]](#) illustrates these planes on a goat (a four-legged animal) and a human being.



Shown are the planes of a quadruped goat and a bipedal human. The midsagittal plane divides the body exactly in half, into right and left portions. The frontal plane divides the front and back, and the transverse plane divides the body into upper and lower portions.

Vertebrate animals have a number of defined body cavities, as illustrated in [\[link\]](#). Two of these are major cavities that contain smaller cavities within them. The **dorsal cavity** contains the cranial and the vertebral (or spinal) cavities. The **ventral cavity** contains the thoracic cavity, which in turn contains the pleural cavity around the lungs and the pericardial cavity, which surrounds the heart. The ventral cavity also contains the abdominopelvic cavity, which can be separated into the abdominal and the pelvic cavities.



Vertebrate animals have two major body cavities. The dorsal cavity, indicated in green, contains the cranial and the spinal cavity. The ventral cavity, indicated in yellow, contains the thoracic cavity and the abdominopelvic cavity. The thoracic cavity is separated from the abdominopelvic cavity by the diaphragm. The thoracic cavity is separated into the abdominal cavity and the pelvic cavity by an imaginary line parallel to the pelvis bones. (credit: modification of work by NCI)

**Note:****Career Connections****Physical Anthropologist**

Physical anthropologists study the adaptation, variability, and evolution of human beings, plus their living and fossil relatives. They can work in a variety of settings, although most will have an academic appointment at a university, usually in an anthropology department or a biology, genetics, or zoology department.

Non-academic positions are available in the automotive and aerospace industries where the focus is on human size, shape, and anatomy. Research by these professionals might range from studies of how the human body reacts to car crashes to exploring how to make seats more comfortable. Other non-academic positions can be obtained in museums of natural history, anthropology, archaeology, or science and technology. These positions involve educating students from grade school through graduate school. Physical anthropologists serve as education coordinators, collection managers, writers for museum publications, and as administrators. Zoos employ these professionals, especially if they have an expertise in primate biology; they work in collection management and captive breeding programs for endangered species. Forensic science utilizes physical anthropology expertise in identifying human and animal remains, assisting in determining the cause of death, and for expert testimony in trials.

**Section Summary**

Animal bodies come in a variety of sizes and shapes. Limits on animal size and shape include impacts to their movement. Diffusion affects their size and development. Bioenergetics describes how animals use and obtain energy in relation to their body size, activity level, and environment.

**Review Questions**



**Exercise:**

**Problem:**

Which type of animal maintains a constant internal body temperature?

- a. endotherm
- b. ectotherm
- c. coelomate
- d. mesoderm

---

**Solution:**

A

**Exercise:**

**Problem:**

The symmetry found in animals that move swiftly is \_\_\_\_\_.

- a. radial
- b. bilateral
- c. sequential
- d. interrupted

---

**Solution:**

B

**Exercise:**

**Problem:**

What term describes the condition of a desert mouse that lowers its metabolic rate and “sleeps” during the hot day?

- a. turgid
- b. hibernation

- c. estivation
- d. normal sleep pattern

---

**Solution:**

C

**Exercise:**

**Problem:**

A plane that divides an animal into equal right and left portions is \_\_\_\_\_.

- a. diagonal
- b. midsagittal
- c. coronal
- d. transverse

---

**Solution:**

B

**Exercise:**

**Problem:**

A plane that divides an animal into dorsal and ventral portions is \_\_\_\_\_.

- a. sagittal
- b. midsagittal
- c. coronal
- d. transverse

---

**Solution:**

D

**Exercise:**

**Problem:** The pleural cavity is a part of which cavity?

- a. dorsal cavity
- b. thoracic cavity
- c. abdominal cavity
- d. pericardial cavity

---

**Solution:**

B

**Free Response****Exercise:****Problem:**

How does diffusion limit the size of an organism? How is this counteracted?

---

**Solution:**

Diffusion is effective over a very short distance. If a cell exceeds this distance in its size, the center of the cell cannot get adequate nutrients nor can it expel enough waste to survive. To compensate for this, cells can loosely adhere to each other in a liquid medium, or develop into multi-celled organisms that use circulatory and respiratory systems to deliver nutrients and remove wastes.

**Exercise:**

**Problem:** What is the relationship between BMR and body size? Why?

---

**Solution:**

Basal Metabolic Rate is an expression of the metabolic processes that occur to maintain an individual's functioning and body temperature. Smaller bodied animals have a relatively large surface area compared to a much larger animal. The small animal's large surface area leads to increased heat loss that the animal must compensate for, resulting in a higher BMR. A large animal, having less relative surface area, does not lose as much heat and has a correspondingly lower BMR.

## **Glossary**

apodeme

ingrowth of an animal's exoskeleton that functions as an attachment site for muscles

asymmetrical

describes animals with no axis of symmetry in their body pattern

basal metabolic rate (BMR)

metabolic rate at rest in endothermic animals

dorsal cavity

body cavity on the posterior or back portion of an animal; includes the cranial and vertebral cavities

ectotherm

animal incapable of maintaining a relatively constant internal body temperature

endotherm

animal capable of maintaining a relatively constant internal body temperature

estivation

torpor in response to extremely high temperatures and low water availability

frontal (coronal) plane

plane cutting through an animal separating the individual into front and back portions

fusiform

animal body shape that is tubular and tapered at both ends

hibernation

torpor over a long period of time, such as a winter

midsagittal plane

plane cutting through an animal separating the individual into even right and left sides

sagittal plane

plane cutting through an animal separating the individual into right and left sides

standard metabolic rate (SMR)

metabolic rate at rest in ectothermic animals

torpor

decrease in activity and metabolism that allows an animal to survive adverse conditions

transverse (horizontal) plane

plane cutting through an animal separating the individual into upper and lower portions

ventral cavity

body cavity on the anterior or front portion of an animal that includes the thoracic cavities and the abdominopelvic cavities

## Animal Primary Tissues

By the end of this section, you will be able to:

- Describe epithelial tissues
- Discuss the different types of connective tissues in animals
- Describe three types of muscle tissues
- Describe nervous tissue

The tissues of multicellular, complex animals are four primary types: epithelial, connective, muscle, and nervous. Recall that tissues are groups of similar cells group of similar cells carrying out related functions. These tissues combine to form organs—like the skin or kidney—that have specific, specialized functions within the body. Organs are organized into organ systems to perform functions; examples include the circulatory system, which consists of the heart and blood vessels, and the digestive system, consisting of several organs, including the stomach, intestines, liver, and pancreas. Organ systems come together to create an entire organism.

## Epithelial Tissues

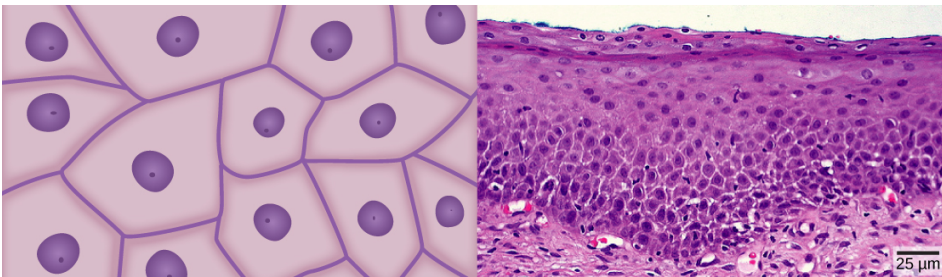
**Epithelial tissues** cover the outside of organs and structures in the body and line the lumens of organs in a single layer or multiple layers of cells. The types of epithelia are classified by the shapes of cells present and the number of layers of cells. Epithelia composed of a single layer of cells is called **simple epithelia**; epithelial tissue composed of multiple layers is called **stratified epithelia**. [\[link\]](#) summarizes the different types of epithelial tissues.

Different Types of Epithelial Tissues		
Cell shape	Description	Location

Different Types of Epithelial Tissues		
Cell shape	Description	Location
squamous	flat, irregular round shape	simple: lung alveoli, capillaries stratified: skin, mouth, vagina
cuboidal	cube shaped, central nucleus	glands, renal tubules
columnar	tall, narrow, nucleus toward base tall, narrow, nucleus along cell	simple: digestive tract pseudostratified: respiratory tract
transitional	round, simple but appear stratified	urinary bladder

## Squamous Epithelia

**Squamous epithelial** cells are generally round, flat, and have a small, centrally located nucleus. The cell outline is slightly irregular, and cells fit together to form a covering or lining. When the cells are arranged in a single layer (simple epithelia), they facilitate diffusion in tissues, such as the areas of gas exchange in the lungs and the exchange of nutrients and waste at blood capillaries.



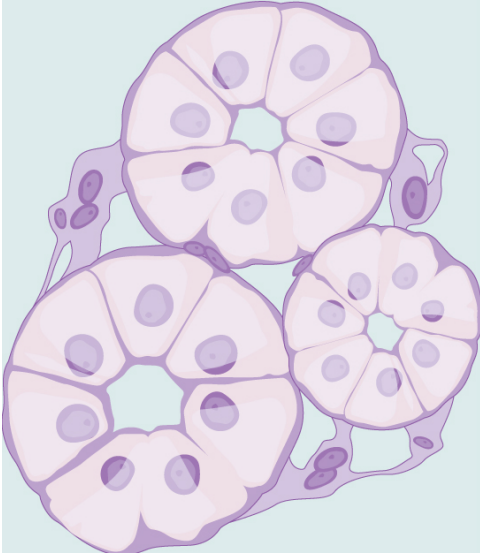
Squamous epithelia cells (a) have a slightly irregular shape, and a small, centrally located nucleus. These cells can be stratified into layers, as in (b) this human cervix specimen. (credit b: modification of work by Ed Uthman; scale-bar data from Matt Russell)

[\[link\]](#)**a** illustrates a layer of squamous cells with their membranes joined together to form an epithelium. Image [\[link\]](#)**b** illustrates squamous epithelial cells arranged in stratified layers, where protection is needed on the body from outside abrasion and damage. This is called a stratified squamous epithelium and occurs in the skin and in tissues lining the mouth and vagina.

## **Cuboidal Epithelia**

**Cuboidal epithelial** cells, shown in [\[link\]](#), are cube-shaped with a single, central nucleus. They are most commonly found in a single layer representing a simple epithelia in glandular tissues throughout the body where they prepare and secrete glandular material. They are also found in the walls of tubules and in the ducts of the kidney and liver.

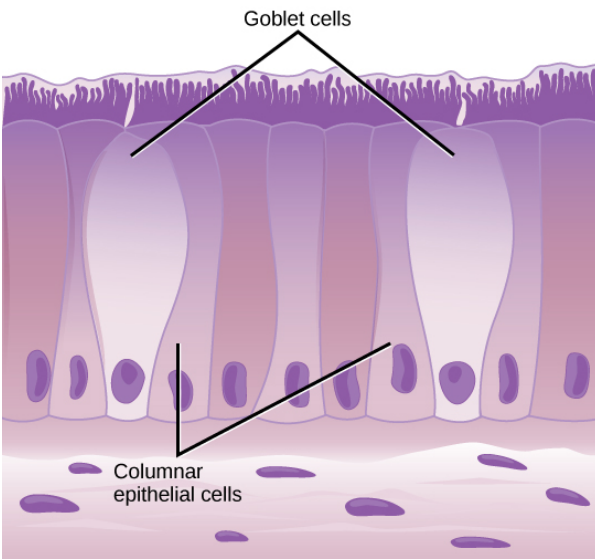




Simple cuboidal epithelial cells line tubules in the mammalian kidney, where they are involved in filtering the blood.

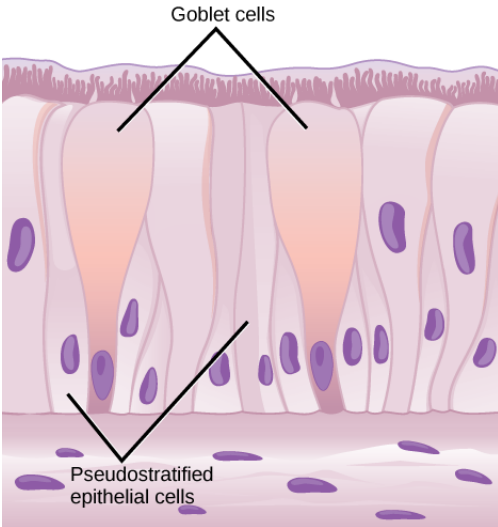
## Columnar Epithelia

**Columnar epithelial** cells are taller than they are wide: they resemble a stack of columns in an epithelial layer, and are most commonly found in a single-layer arrangement. The nuclei of columnar epithelial cells in the digestive tract appear to be lined up at the base of the cells, as illustrated in [\[link\]](#). These cells absorb material from the lumen of the digestive tract and prepare it for entry into the body through the circulatory and lymphatic systems.



Simple columnar epithelial cells absorb material from the digestive tract. Goblet cells secrete mucous into the digestive tract lumen.

Columnar epithelial cells lining the respiratory tract appear to be stratified. However, each cell is attached to the base membrane of the tissue and, therefore, they are simple tissues. The nuclei are arranged at different levels in the layer of cells, making it appear as though there is more than one layer, as seen in [\[link\]](#). This is called **pseudostratified**, columnar epithelia. This cellular covering has cilia at the apical, or free, surface of the cells. The cilia enhance the movement of mucous and trapped particles out of the respiratory tract, helping to protect the system from invasive microorganisms and harmful material that has been breathed into the body. Goblet cells are interspersed in some tissues (such as the lining of the trachea). The goblet cells contain mucous that traps irritants, which in the case of the trachea keep these irritants from getting into the lungs.



Pseudostratified columnar epithelia line the respiratory tract. They exist in one layer, but the arrangement of nuclei at different levels makes it appear that there is more than one layer. Goblet cells interspersed between the columnar epithelial cells secrete mucous into the respiratory tract.

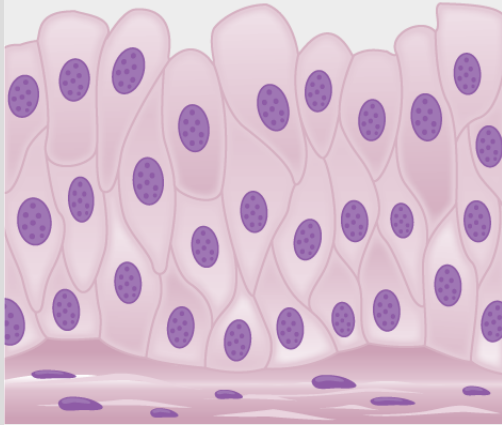
## Transitional Epithelia

**Transitional** or uroepithelial cells appear only in the urinary system, primarily in the bladder and ureter. These cells are arranged in a stratified layer, but they have the capability of appearing to pile up on top of each other in a relaxed, empty bladder, as illustrated in [\[link\]](#). As the urinary bladder fills, the epithelial layer unfolds and expands to hold the volume of

urine introduced into it. As the bladder fills, it expands and the lining becomes thinner. In other words, the tissue transitions from thick to thin.

**Note:**

**Art Connection**



Transitional epithelia of the urinary bladder undergo changes in thickness depending on how full the bladder is.

Which of the following statements about types of epithelial cells is false?

- a. Simple columnar epithelial cells line the tissue of the lung.
- b. Simple cuboidal epithelial cells are involved in the filtering of blood in the kidney.
- c. Pseudostratified columnar epithelia occur in a single layer, but the arrangement of nuclei makes it appear that more than one layer is present.
- d. Transitional epithelia change in thickness depending on how full the bladder is.

# Connective Tissues

**Connective tissues** are made up of a matrix consisting of living cells and a non-living substance, called the ground substance. The ground substance is made of an organic substance (usually a protein) and an inorganic substance (usually a mineral or water). The principal cell of connective tissues is the fibroblast. This cell makes the fibers found in nearly all of the connective tissues. Fibroblasts are motile, able to carry out mitosis, and can synthesize whichever connective tissue is needed. Macrophages, lymphocytes, and, occasionally, leukocytes can be found in some of the tissues. Some tissues have specialized cells that are not found in the others. The **matrix** in connective tissues gives the tissue its density. When a connective tissue has a high concentration of cells or fibers, it has proportionally a less dense matrix.

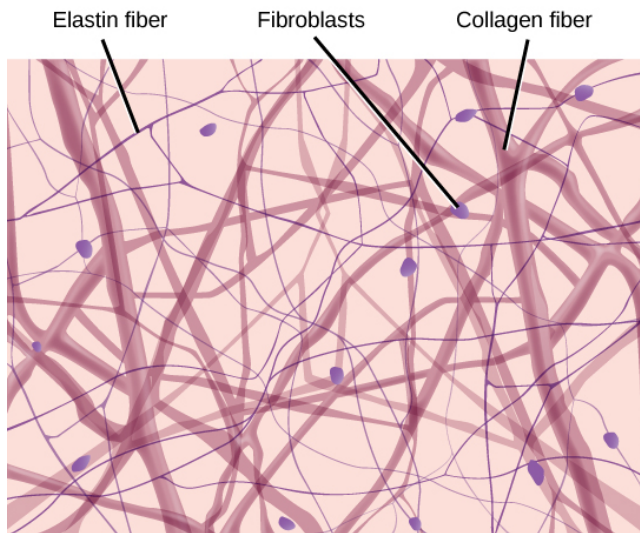
The organic portion or protein fibers found in connective tissues are either collagen, elastic, or reticular fibers. Collagen fibers provide strength to the tissue, preventing it from being torn or separated from the surrounding tissues. Elastic fibers are made of the protein elastin; this fiber can stretch to one and one half of its length and return to its original size and shape. Elastic fibers provide flexibility to the tissues. Reticular fibers are the third type of protein fiber found in connective tissues. This fiber consists of thin strands of collagen that form a network of fibers to support the tissue and other organs to which it is connected. The various types of connective tissues, the types of cells and fibers they are made of, and sample locations of the tissues is summarized in [\[link\]](#).

Connective Tissues			
Tissue	Cells	Fibers	Location

<b>Connective Tissues</b>			
<b>Tissue</b>	<b>Cells</b>	<b>Fibers</b>	<b>Location</b>
loose/areolar	fibroblasts, macrophages, some lymphocytes, some neutrophils	few: collagen, elastic, reticular	around blood vessels; anchors epithelia
dense, fibrous connective tissue	fibroblasts, macrophages,	mostly collagen	irregular: skin regular: tendons, ligaments
cartilage	chondrocytes, chondroblasts	hyaline: few collagen fibrocartilage: large amount of collagen	shark skeleton, fetal bones, human ears, intervertebral discs
bone	osteoblasts, osteocytes, osteoclasts	some: collagen, elastic	vertebrate skeletons
adipose	adipocytes	few	adipose (fat)
blood	red blood cells, white blood cells	none	blood

### **Loose/Areolar Connective Tissue**

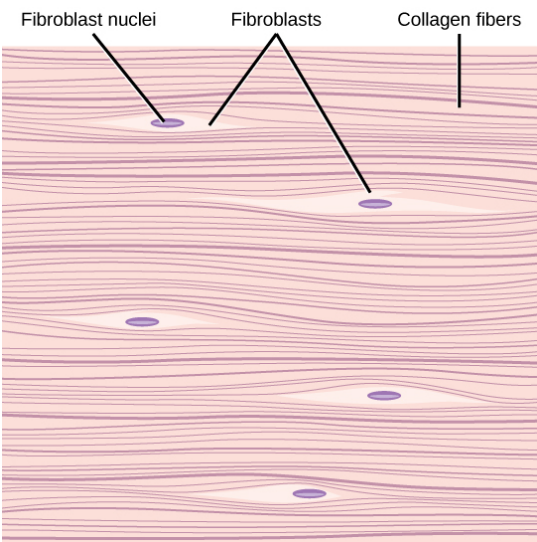
**Loose connective tissue**, also called areolar connective tissue, has a sampling of all of the components of a connective tissue. As illustrated in [\[link\]](#), loose connective tissue has some fibroblasts; macrophages are present as well. Collagen fibers are relatively wide and stain a light pink, while elastic fibers are thin and stain dark blue to black. The space between the formed elements of the tissue is filled with the matrix. The material in the connective tissue gives it a loose consistency similar to a cotton ball that has been pulled apart. Loose connective tissue is found around every blood vessel and helps to keep the vessel in place. The tissue is also found around and between most body organs. In summary, areolar tissue is tough, yet flexible, and comprises membranes.



Loose connective tissue is composed of loosely woven collagen and elastic fibers. The fibers and other components of the connective tissue matrix are secreted by fibroblasts.

## Fibrous Connective Tissue

**Fibrous connective tissues** contain large amounts of collagen fibers and few cells or matrix material. The fibers can be arranged irregularly or regularly with the strands lined up in parallel. Irregularly arranged fibrous connective tissues are found in areas of the body where stress occurs from all directions, such as the dermis of the skin. Regular fibrous connective tissue, shown in [\[link\]](#), is found in tendons (which connect muscles to bones) and ligaments (which connect bones to bones).



Fibrous connective tissue from the tendon has strands of collagen fibers lined up in parallel.

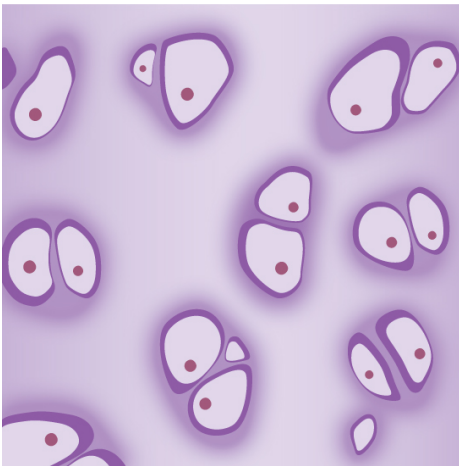
## Cartilage

**Cartilage** is a connective tissue with a large amount of the matrix and variable amounts of fibers. The cells, called **chondrocytes**, make the matrix



and fibers of the tissue. Chondrocytes are found in spaces within the tissue called **lacunae**.

A cartilage with few collagen and elastic fibers is hyaline cartilage, illustrated in [\[link\]](#). The lacunae are randomly scattered throughout the tissue and the matrix takes on a milky or scrubbed appearance with routine histological stains. Sharks have cartilaginous skeletons, as does nearly the entire human skeleton during a specific pre-birth developmental stage. A remnant of this cartilage persists in the outer portion of the human nose. Hyaline cartilage is also found at the ends of long bones, reducing friction and cushioning the articulations of these bones.



Hyaline cartilage consists of a matrix with cells called chondrocytes embedded in it. The chondrocytes exist in cavities in the matrix called lacunae.

Elastic cartilage has a large amount of elastic fibers, giving it tremendous flexibility. The ears of most vertebrate animals contain this cartilage as do portions of the larynx, or voice box. Fibrocartilage contains a large amount of collagen fibers, giving the tissue tremendous strength. Fibrocartilage comprises the intervertebral discs in vertebrate animals. Hyaline cartilage found in movable joints such as the knee and shoulder becomes damaged as a result of age or trauma. Damaged hyaline cartilage is replaced by fibrocartilage and results in the joints becoming “stiff.”

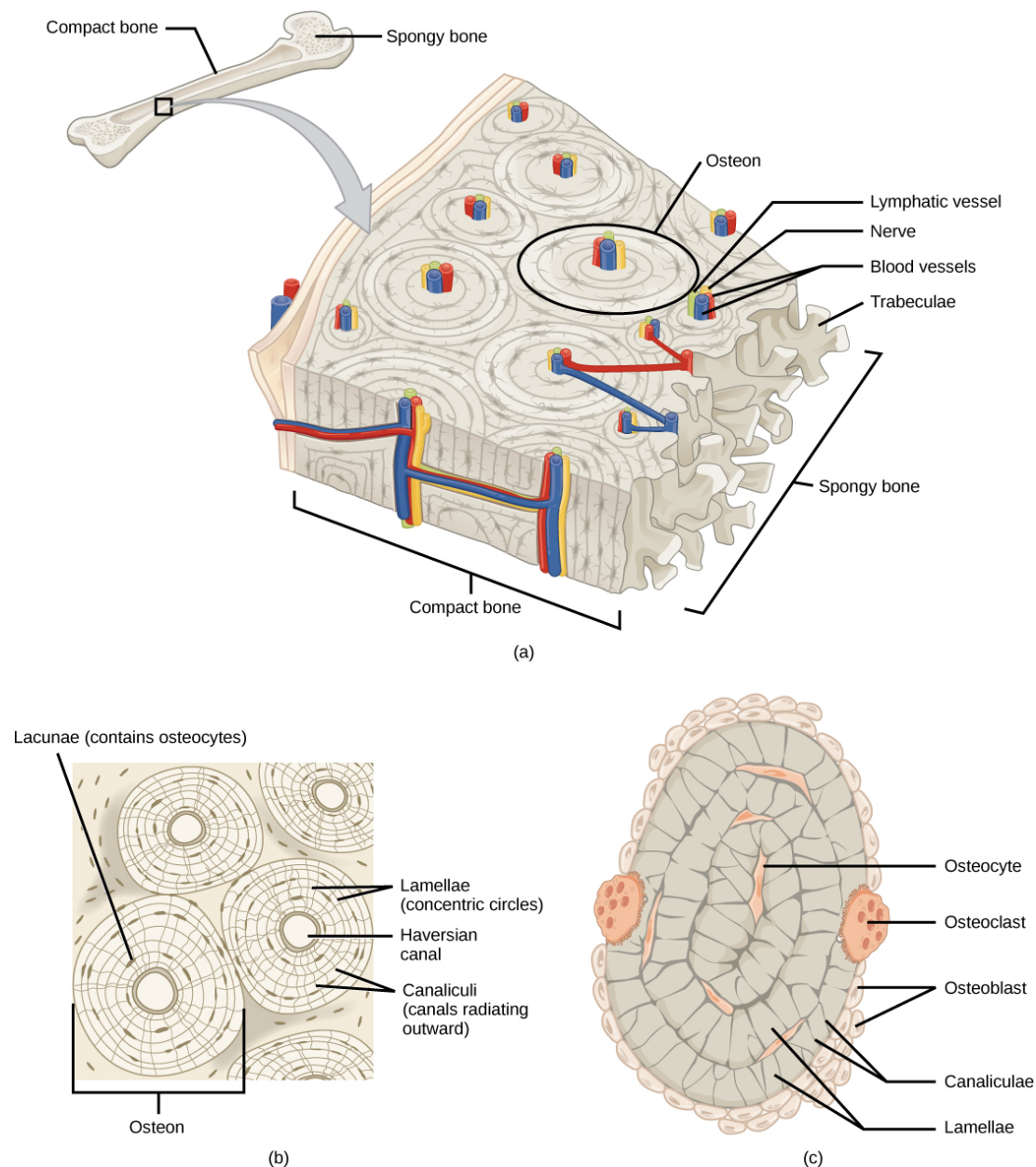
## Bone

Bone, or osseous tissue, is a connective tissue that has a large amount of two different types of matrix material. The organic matrix is similar to the matrix material found in other connective tissues, including some amount of collagen and elastic fibers. This gives strength and flexibility to the tissue. The inorganic matrix consists of mineral salts—mostly calcium salts—that give the tissue hardness. Without adequate organic material in the matrix, the tissue breaks; without adequate inorganic material in the matrix, the tissue bends.

There are three types of cells in bone: osteoblasts, osteocytes, and osteoclasts. Osteoblasts are active in making bone for growth and remodeling. Osteoblasts deposit bone material into the matrix and, after the matrix surrounds them, they continue to live, but in a reduced metabolic state as osteocytes. Osteocytes are found in lacunae of the bone. Osteoclasts are active in breaking down bone for bone remodeling, and they provide access to calcium stored in tissues. Osteoclasts are usually found on the surface of the tissue.

Bone can be divided into two types: compact and spongy. Compact bone is found in the shaft (or diaphysis) of a long bone and the surface of the flat bones, while spongy bone is found in the end (or epiphysis) of a long bone. Compact bone is organized into subunits called **osteons**, as illustrated in [\[link\]](#). A blood vessel and a nerve are found in the center of the structure within the Haversian canal, with radiating circles of lacunae around it known as lamellae. The wavy lines seen between the lacunae are

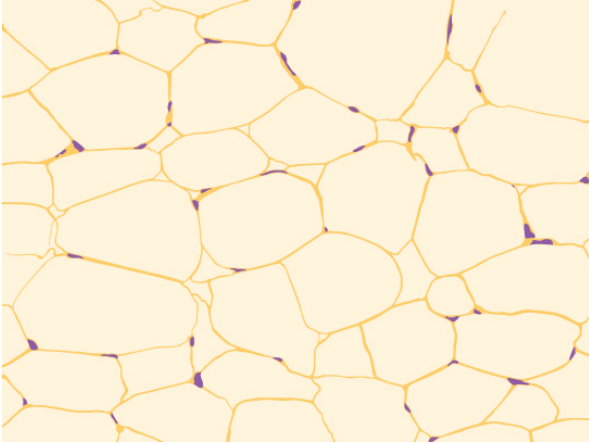
microchannels called **canaliculi**; they connect the lacunae to aid diffusion between the cells. Spongy bone is made of tiny plates called **trabeculae** these plates serve as struts to give the spongy bone strength. Over time, these plates can break causing the bone to become less resilient. Bone tissue forms the internal skeleton of vertebrate animals, providing structure to the animal and points of attachment for tendons.



(a) Compact bone is a dense matrix on the outer surface of bone. Spongy bone, inside the compact bone, is porous with web-like trabeculae. (b) Compact bone is organized into rings called osteons. Blood vessels, nerves, and lymphatic vessels are found in the central Haversian canal. Rings of lamellae surround the Haversian canal. Between the lamellae are cavities called lacunae. Canaliculi are microchannels connecting the lacunae together. (c) Osteoblasts surround the exterior of the bone. Osteoclasts bore tunnels into the bone and osteocytes are found in the lacunae.

## **Adipose Tissue**

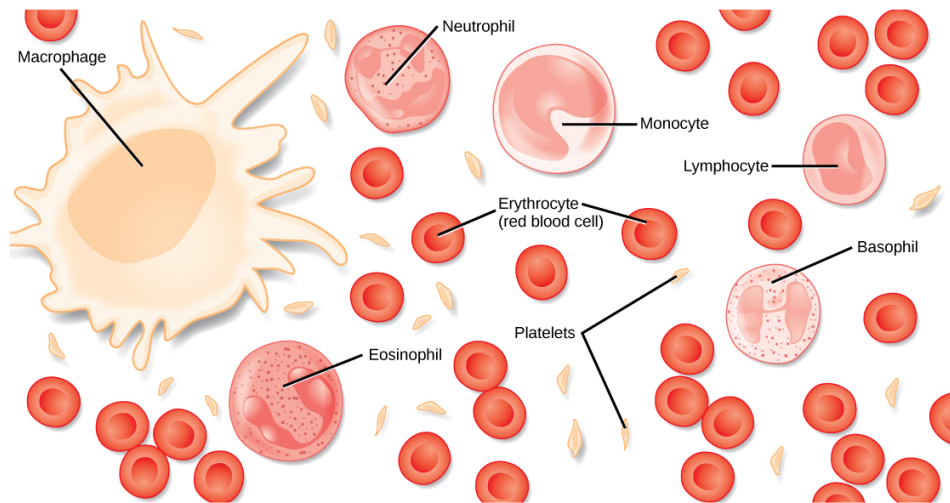
Adipose tissue, or fat tissue, is considered a connective tissue even though it does not have fibroblasts or a real matrix and only has a few fibers. Adipose tissue is made up of cells called adipocytes that collect and store fat in the form of triglycerides, for energy metabolism. Adipose tissues additionally serve as insulation to help maintain body temperatures, allowing animals to be endothermic, and they function as cushioning against damage to body organs. Under a microscope, adipose tissue cells appear empty due to the extraction of fat during the processing of the material for viewing, as seen in [\[link\]](#). The thin lines in the image are the cell membranes, and the nuclei are the small, black dots at the edges of the cells.



Adipose is a connective tissue  
is made up of cells called  
adipocytes. Adipocytes have  
small nuclei localized at the  
cell edge.

## **Blood**

Blood is considered a connective tissue because it has a matrix, as shown in [\[link\]](#). The living cell types are red blood cells (RBC), also called erythrocytes, and white blood cells (WBC), also called leukocytes. The fluid portion of whole blood, its matrix, is commonly called plasma.



Blood is a connective tissue that has a fluid matrix, called plasma, and no fibers. Erythrocytes (red blood cells), the predominant cell type, are involved in the transport of oxygen and carbon dioxide. Also present are various leukocytes (white blood cells) involved in immune response.

The cell found in greatest abundance in blood is the erythrocyte. Erythrocytes are counted in millions in a blood sample: the average number of red blood cells in primates is 4.7 to 5.5 million cells per microliter. Erythrocytes are consistently the same size in a species, but vary in size between species. For example, the average diameter of a primate red blood cell is 7.5  $\mu\text{l}$ , a dog is close at 7.0  $\mu\text{l}$ , but a cat's RBC diameter is 5.9  $\mu\text{l}$ . Sheep erythrocytes are even smaller at 4.6  $\mu\text{l}$ . Mammalian erythrocytes lose their nuclei and mitochondria when they are released from the bone marrow where they are made. Fish, amphibian, and avian red blood cells maintain their nuclei and mitochondria throughout the cell's life. The principal job of an erythrocyte is to carry and deliver oxygen to the tissues.

Leukocytes are the predominant white blood cells found in the peripheral blood. Leukocytes are counted in the thousands in the blood with measurements expressed as ranges: primate counts range from 4,800 to 10,800 cells per  $\mu\text{l}$ , dogs from 5,600 to 19,200 cells per  $\mu\text{l}$ , cats from 8,000

to 25,000 cells per  $\mu\text{l}$ , cattle from 4,000 to 12,000 cells per  $\mu\text{l}$ , and pigs from 11,000 to 22,000 cells per  $\mu\text{l}$ .

Lymphocytes function primarily in the immune response to foreign antigens or material. Different types of lymphocytes make antibodies tailored to the foreign antigens and control the production of those antibodies. Neutrophils are phagocytic cells and they participate in one of the early lines of defense against microbial invaders, aiding in the removal of bacteria that has entered the body. Another leukocyte that is found in the peripheral blood is the monocyte. Monocytes give rise to phagocytic macrophages that clean up dead and damaged cells in the body, whether they are foreign or from the host animal. Two additional leukocytes in the blood are eosinophils and basophils—both help to facilitate the inflammatory response.

The slightly granular material among the cells is a cytoplasmic fragment of a cell in the bone marrow. This is called a platelet or thrombocyte. Platelets participate in the stages leading up to coagulation of the blood to stop bleeding through damaged blood vessels. Blood has a number of functions, but primarily it transports material through the body to bring nutrients to cells and remove waste material from them.

## **Muscle Tissues**

There are three types of muscle in animal bodies: smooth, skeletal, and cardiac. They differ by the presence or absence of striations or bands, the number and location of nuclei, whether they are voluntarily or involuntarily controlled, and their location within the body. [\[link\]](#) summarizes these differences.

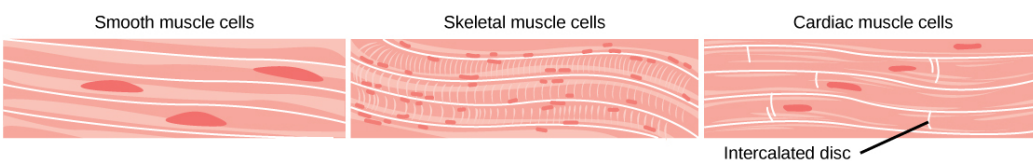
### **Types of Muscles**

---

<b>Types of Muscles of Muscle</b>				
	<b>Striations</b>	<b>Nuclei</b>	<b>Control</b>	<b>Location</b>
<b>Type of Muscle</b>	<b>Striations</b>	<b>Nuclei</b>	<b>Control</b>	<b>Location</b>
smooth	no	single, in center	involuntary	visceral organs
skeletal	yes	many, at periphery	voluntary	skeletal muscles
cardiac	yes	single, in center	involuntary	heart

## Smooth Muscle

Smooth muscle does not have striations in its cells. It has a single, centrally located nucleus, as shown in [\[link\]](#). Constriction of smooth muscle occurs under involuntary, autonomic nervous control and in response to local conditions in the tissues. Smooth muscle tissue is also called non-striated as it lacks the banded appearance of skeletal and cardiac muscle. The walls of blood vessels, the tubes of the digestive system, and the tubes of the reproductive systems are composed of mostly smooth muscle.



Smooth muscle cells do not have striations, while skeletal



muscle cells do. Cardiac muscle cells have striations, but, unlike the multinucleate skeletal cells, they have only one nucleus. Cardiac muscle tissue also has intercalated discs, specialized regions running along the plasma membrane that join adjacent cardiac muscle cells and assist in passing an electrical impulse from cell to cell.

## **Skeletal Muscle**

Skeletal muscle has striations across its cells caused by the arrangement of the contractile proteins actin and myosin. These muscle cells are relatively long and have multiple nuclei along the edge of the cell. Skeletal muscle is under voluntary, somatic nervous system control and is found in the muscles that move bones. [\[link\]](#) illustrates the histology of skeletal muscle.

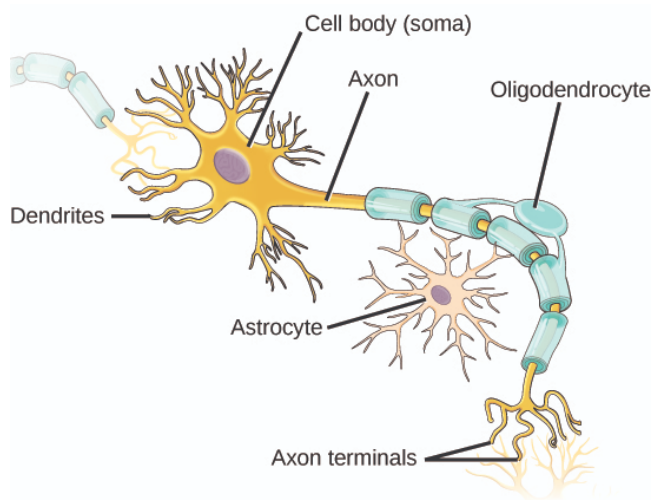
## **Cardiac Muscle**

Cardiac muscle, shown in [\[link\]](#), is found only in the heart. Like skeletal muscle, it has cross striations in its cells, but cardiac muscle has a single, centrally located nucleus. Cardiac muscle is not under voluntary control but can be influenced by the autonomic nervous system to speed up or slow down. An added feature to cardiac muscle cells is a line that extends along the end of the cell as it abuts the next cardiac cell in the row. This line is called an intercalated disc: it assists in passing electrical impulse efficiently from one cell to the next and maintains the strong connection between neighboring cardiac cells.

## **Nervous Tissues**

Nervous tissues are made of cells specialized to receive and transmit electrical impulses from specific areas of the body and to send them to specific locations in the body. The main cell of the nervous system is the

neuron, illustrated in [\[link\]](#). The large structure with a central nucleus is the cell body of the neuron. Projections from the cell body are either dendrites specialized in receiving input or a single axon specialized in transmitting impulses. Some glial cells are also shown. Astrocytes regulate the chemical environment of the nerve cell, and oligodendrocytes insulate the axon so the electrical nerve impulse is transferred more efficiently. Other glial cells that are not shown support the nutritional and waste requirements of the neuron. Some of the glial cells are phagocytic and remove debris or damaged cells from the tissue. A nerve consists of neurons and glial cells.



The neuron has projections called dendrites that receive signals and projections called axons that send signals. Also shown are two types of glial cells: astrocytes regulate the chemical environment of the nerve cell, and oligodendrocytes insulate the axon so the electrical nerve impulse is transferred more efficiently.

**Note:**

Link to Learning



Click through the [interactive review](#) to learn more about epithelial tissues.

**Note:**

Career Connections

**Pathologist**

A pathologist is a medical doctor or veterinarian who has specialized in the laboratory detection of disease in animals, including humans. These professionals complete medical school education and follow it with an extensive post-graduate residency at a medical center. A pathologist may oversee clinical laboratories for the evaluation of body tissue and blood samples for the detection of disease or infection. They examine tissue specimens through a microscope to identify cancers and other diseases. Some pathologists perform autopsies to determine the cause of death and the progression of disease.

## Section Summary

The basic building blocks of complex animals are four primary tissues. These are combined to form organs, which have a specific, specialized function within the body, such as the skin or kidney. Organs are organized together to perform common functions in the form of systems. The four primary tissues are epithelia, connective tissues, muscle tissues, and nervous tissues.

## Art Connections

### Exercise:

#### Problem:

[\[link\]](#) Which of the following statements about types of epithelial cells is false?

- a. Simple columnar epithelial cells line the tissue of the lung.
- b. Simple cuboidal epithelial cells are involved in the filtering of blood in the kidney.
- c. Pseudostratified columnar epithelia occur in a single layer, but the arrangement of nuclei makes it appear that more than one layer is present.
- d. Transitional epithelia change in thickness depending on how full the bladder is.

---

#### Solution:

[\[link\]](#) A

## Review Questions

### Exercise:

**Problem:** Which type of epithelial cell is best adapted to aid diffusion?

- a. squamous
- b. cuboidal
- c. columnar
- d. transitional

---

#### Solution:

A

**Exercise:**

**Problem:** Which type of epithelial cell is found in glands?

- a. squamous
  - b. cuboidal
  - c. columnar
  - d. transitional
- 

**Solution:**

B

**Exercise:**

**Problem:** Which type of epithelial cell is found in the urinary bladder?

- a. squamous
  - b. cuboidal
  - c. columnar
  - d. transitional
- 

**Solution:**

D

**Exercise:**

**Problem:** Which type of connective tissue has the most fibers?

- a. loose connective tissue
- b. fibrous connective tissue
- c. cartilage
- d. bone

---

**Solution:**

B

**Exercise:**

**Problem:**

Which type of connective tissue has a mineralized different matrix?

- a. loose connective tissue
- b. fibrous connective tissue
- c. cartilage
- d. bone

---

**Solution:**

D

**Exercise:**

**Problem:**

The cell found in bone that breaks it down is called an \_\_\_\_\_.

- a. osteoblast
- b. osteocyte
- c. osteoclast
- d. osteon

---

**Solution:**

C

**Exercise:**

**Problem:**

The cell found in bone that makes the bone is called an \_\_\_\_\_.

- a. osteoblast
- b. osteocyte
- c. osteoclast
- d. osteon

---

**Solution:**

A

**Exercise:**

**Problem:** Plasma is the \_\_\_\_\_.

- a. fibers in blood
- b. matrix of blood
- c. cell that phagocytizes bacteria
- d. cell fragment found in the tissue

---

**Solution:**

B

**Exercise:**

**Problem:**

The type of muscle cell under voluntary control is the \_\_\_\_\_.

- a. smooth muscle
- b. skeletal muscle
- c. cardiac muscle
- d. visceral muscle

---

**Solution:**

B

**Exercise:**

**Problem:** The part of a neuron that contains the nucleus is the

- a. cell body
- b. dendrite
- c. axon
- d. glial

---

**Solution:**

A

**Free Response****Exercise:****Problem:**

How can squamous epithelia both facilitate diffusion and prevent damage from abrasion?

---

**Solution:**

Squamous epithelia can be either simple or stratified. As a single layer of cells, it presents a very thin epithelia that minimally inhibits diffusion. As a stratified epithelia, the surface cells can be sloughed off and the cells in deeper layers protect the underlying tissues from damage.

**Exercise:**

**Problem:** What are the similarities between cartilage and bone?

---

**Solution:**



Both contain cells other than the traditional fibroblast. Both have cells that lodge in spaces within the tissue called lacunae. Both collagen and elastic fibers are found in bone and cartilage. Both tissues participate in vertebrate skeletal development and formation.

## **Glossary**

### canaliculus

microchannel that connects the lacunae and aids diffusion between cells

### cartilage

type of connective tissue with a large amount of ground substance matrix, cells called chondrocytes, and some amount of fibers

### chondrocyte

cell found in cartilage

### columnar epithelia

epithelia made of cells taller than they are wide, specialized in absorption

### connective tissue

type of tissue made of cells, ground substance matrix, and fibers

### cuboidal epithelia

epithelia made of cube-shaped cells, specialized in glandular functions

### epithelial tissue

tissue that either lines or covers organs or other tissues

### fibrous connective tissue

type of connective tissue with a high concentration of fibers

### lacuna

space in cartilage and bone that contains living cells

loose (areolar) connective tissue

type of connective tissue with small amounts of cells, matrix, and fibers; found around blood vessels

matrix

component of connective tissue made of both living and non-living (ground substances) cells

osteon

subunit of compact bone

pseudostratified

layer of epithelia that appears multilayered, but is a simple covering

simple epithelia

single layer of epithelial cells

squamous epithelia

type of epithelia made of flat cells, specialized in aiding diffusion or preventing abrasion

stratified epithelia

multiple layers of epithelial cells

trabecula

tiny plate that makes up spongy bone and gives it strength

transitional epithelia

epithelia that can transition for appearing multilayered to simple; also called uroepithelial

## Homeostasis

By the end of this section, you will be able to:

- Define homeostasis
- Describe the factors affecting homeostasis
- Discuss positive and negative feedback mechanisms used in homeostasis
- Describe thermoregulation of endothermic and ectothermic animals

Animal organs and organ systems constantly adjust to internal and external changes through a process called homeostasis (“steady state”). These changes might be in the level of glucose or calcium in blood or in external temperatures. **Homeostasis** means to maintain dynamic equilibrium in the body. It is dynamic because it is constantly adjusting to the changes that the body’s systems encounter. It is equilibrium because body functions are kept within specific ranges. Even an animal that is apparently inactive is maintaining this homeostatic equilibrium.

## Homeostatic Process

The goal of homeostasis is the maintenance of equilibrium around a point or value called a **set point**. While there are normal fluctuations from the set point, the body’s systems will usually attempt to go back to this point. A change in the internal or external environment is called a stimulus and is detected by a receptor; the response of the system is to adjust the deviation parameter toward the set point. For instance, if the body becomes too warm, adjustments are made to cool the animal. If the blood’s glucose rises after a meal, adjustments are made to lower the blood glucose level by getting the nutrient into tissues that need it or to store it for later use.

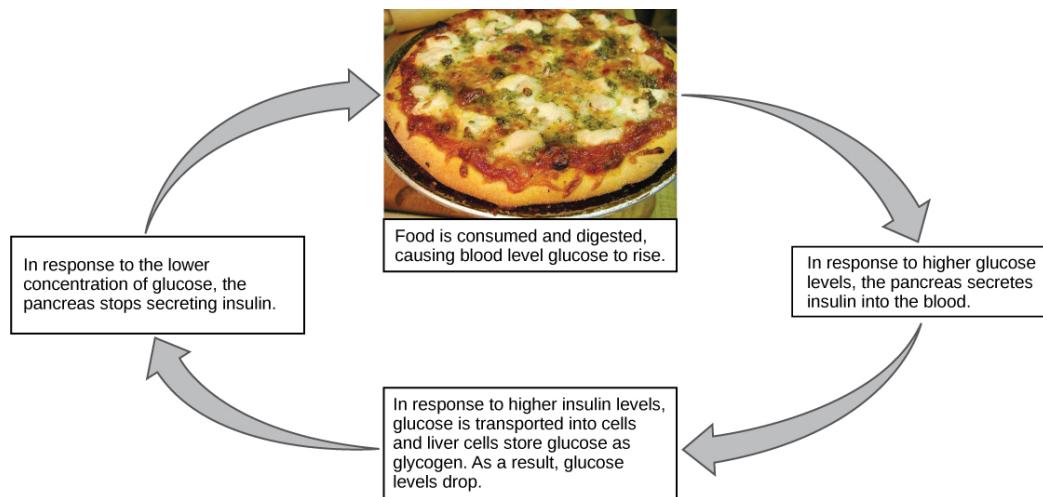
## Control of Homeostasis

When a change occurs in an animal’s environment, an adjustment must be made. The receptor senses the change in the environment, then sends a signal to the control center (in most cases, the brain) which in turn generates a response that is signaled to an effector. The effector is a muscle (that contracts or relaxes) or a gland that secretes. Homeostasis is

maintained by negative feedback loops. Positive feedback loops actually push the organism further out of homeostasis, but may be necessary for life to occur. Homeostasis is controlled by the nervous and endocrine system of mammals.

## **Negative Feedback Mechanisms**

Any homeostatic process that changes the direction of the stimulus is a **negative feedback loop**. It may either increase or decrease the stimulus, but the stimulus is not allowed to continue as it did before the receptor sensed it. In other words, if a level is too high, the body does something to bring it down, and conversely, if a level is too low, the body does something to make it go up. Hence the term negative feedback. An example is animal maintenance of blood glucose levels. When an animal has eaten, blood glucose levels rise. This is sensed by the nervous system. Specialized cells in the pancreas sense this, and the hormone insulin is released by the endocrine system. Insulin causes blood glucose levels to decrease, as would be expected in a negative feedback system, as illustrated in [\[link\]](#). However, if an animal has not eaten and blood glucose levels decrease, this is sensed in another group of cells in the pancreas, and the hormone glucagon is released causing glucose levels to increase. This is still a negative feedback loop, but not in the direction expected by the use of the term “negative.” Another example of an increase as a result of the feedback loop is the control of blood calcium. If calcium levels decrease, specialized cells in the parathyroid gland sense this and release parathyroid hormone (PTH), causing an increased absorption of calcium through the intestines and kidneys and, possibly, the breakdown of bone in order to liberate calcium. The effects of PTH are to raise blood levels of the element. Negative feedback loops are the predominant mechanism used in homeostasis.

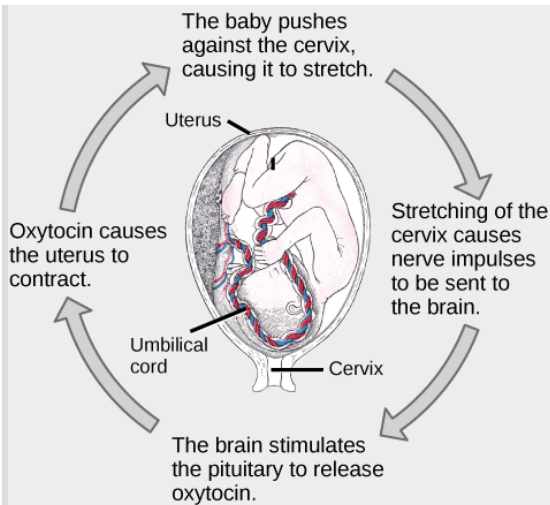


Blood sugar levels are controlled by a negative feedback loop. (credit: modification of work by Jon Sullivan)

## Positive Feedback Loop

A **positive feedback loop** maintains the direction of the stimulus, possibly accelerating it. Few examples of positive feedback loops exist in animal bodies, but one is found in the cascade of chemical reactions that result in blood clotting, or coagulation. As one clotting factor is activated, it activates the next factor in sequence until a fibrin clot is achieved. The direction is maintained, not changed, so this is positive feedback. Another example of positive feedback is uterine contractions during childbirth, as illustrated in [\[link\]](#). The hormone oxytocin, made by the endocrine system, stimulates the contraction of the uterus. This produces pain sensed by the nervous system. Instead of lowering the oxytocin and causing the pain to subside, more oxytocin is produced until the contractions are powerful enough to produce childbirth.

**Note:**  
Art Connection



The birth of a human infant is the result of positive feedback.

State whether each of the following processes is regulated by a positive feedback loop or a negative feedback loop.

- A person feels satiated after eating a large meal.
- The blood has plenty of red blood cells. As a result, erythropoietin, a hormone that stimulates the production of new red blood cells, is no longer released from the kidney.

## Set Point

It is possible to adjust a system's set point. When this happens, the feedback loop works to maintain the new setting. An example of this is blood pressure: over time, the normal or set point for blood pressure can increase as a result of continued increases in blood pressure. The body no longer recognizes the elevation as abnormal and no attempt is made to return to the lower set point. The result is the maintenance of an elevated blood pressure that can have harmful effects on the body. Medication can lower blood

pressure and lower the set point in the system to a more healthy level. This is called a process of **alteration** of the set point in a feedback loop.

Changes can be made in a group of body organ systems in order to maintain a set point in another system. This is called **acclimatization**. This occurs, for instance, when an animal migrates to a higher altitude than it is accustomed to. In order to adjust to the lower oxygen levels at the new altitude, the body increases the number of red blood cells circulating in the blood to ensure adequate oxygen delivery to the tissues. Another example of acclimatization is animals that have seasonal changes in their coats: a heavier coat in the winter ensures adequate heat retention, and a light coat in summer assists in keeping body temperature from rising to harmful levels.

**Note:**

Link to Learning



Feedback mechanisms can be understood in terms of driving a race car along a track: watch a short video lesson on positive and negative feedback loops.

[https://www.openstaxcollege.org/l/feedback\\_loops](https://www.openstaxcollege.org/l/feedback_loops)

## Homeostasis: Thermoregulation

Body temperature affects body activities. Generally, as body temperature rises, enzyme activity rises as well. For every ten degree centigrade rise in temperature, enzyme activity doubles, up to a point. Body proteins, including enzymes, begin to denature and lose their function with high heat

(around 50°C for mammals). Enzyme activity will decrease by half for every ten degree centigrade drop in temperature, to the point of freezing, with a few exceptions. Some fish can withstand freezing solid and return to normal with thawing.

**Note:**

Link to Learning



Watch this Discovery Channel video on thermoregulation to see illustrations of this process in a variety of animals.

<https://www.openstaxcollege.org/l/thermoregulate>

## Endotherms and Ectotherms

Animals can be divided into two groups: some maintain a constant body temperature in the face of differing environmental temperatures, while others have a body temperature that is the same as their environment and thus varies with the environment. Animals that do not control their body temperature are ectotherms. This group has been called cold-blooded, but the term may not apply to an animal in the desert with a very warm body temperature. In contrast to ectotherms, which rely on external temperatures to set their body temperatures, poikilotherms are animals with constantly varying internal temperatures. An animal that maintains a constant body temperature in the face of environmental changes is called a homeotherm. Endotherms are animals that rely on internal sources for body temperature but which can exhibit extremes in temperature. These animals are able to maintain a level of activity at cooler temperature, which an ectotherm cannot due to differing enzyme levels of activity.



Heat can be exchanged between an animal and its environment through four mechanisms: radiation, evaporation, convection, and conduction ([link](#)). Radiation is the emission of electromagnetic “heat” waves. Heat comes from the sun in this manner and radiates from dry skin the same way. Heat can be removed with liquid from a surface during evaporation. This occurs when a mammal sweats. Convection currents of air remove heat from the surface of dry skin as the air passes over it. Heat will be conducted from one surface to another during direct contact with the surfaces, such as an animal resting on a warm rock.



(a) Radiation



(b) Evaporation



(c) Convection



(d) Conduction

Heat can be exchanged by four mechanisms: (a) radiation, (b) evaporation, (c) convection, or (d) conduction. (credit b: modification of work by “Kullez”/Flickr; credit c: modification of work by Chad Rosenthal; credit d: modification of work by “stacey.d”/Flickr)

## Heat Conservation and Dissipation

Animals conserve or dissipate heat in a variety of ways. In certain climates, endothermic animals have some form of insulation, such as fur, fat, feathers, or some combination thereof. Animals with thick fur or feathers create an insulating layer of air between their skin and internal organs. Polar bears and seals live and swim in a subfreezing environment and yet maintain a constant, warm, body temperature. The arctic fox, for example, uses its fluffy tail as extra insulation when it curls up to sleep in cold weather. Mammals have a residual effect from shivering and increased muscle activity: arrector pili muscles cause “goose bumps,” causing small hairs to stand up when the individual is cold; this has the intended effect of increasing body temperature. Mammals use layers of fat to achieve the same end. Loss of significant amounts of body fat will compromise an individual’s ability to conserve heat.

Endotherms use their circulatory systems to help maintain body temperature. Vasodilation brings more blood and heat to the body surface, facilitating radiation and evaporative heat loss, which helps to cool the body. Vasoconstriction reduces blood flow in peripheral blood vessels, forcing blood toward the core and the vital organs found there, and conserving heat. Some animals have adaptations to their circulatory system that enable them to transfer heat from arteries to veins, warming blood returning to the heart. This is called a countercurrent heat exchange; it prevents the cold venous blood from cooling the heart and other internal organs. This adaptation can be shut down in some animals to prevent overheating the internal organs. The countercurrent adaptation is found in many animals, including dolphins, sharks, bony fish, bees, and hummingbirds. In contrast, similar adaptations can help cool endotherms when needed, such as dolphin flukes and elephant ears.

Some ectothermic animals use changes in their behavior to help regulate body temperature. For example, a desert ectothermic animal may simply seek cooler areas during the hottest part of the day in the desert to keep from getting too warm. The same animals may climb onto rocks to capture heat during a cold desert night. Some animals seek water to aid evaporation

in cooling them, as seen with reptiles. Other ectotherms use group activity such as the activity of bees to warm a hive to survive winter.

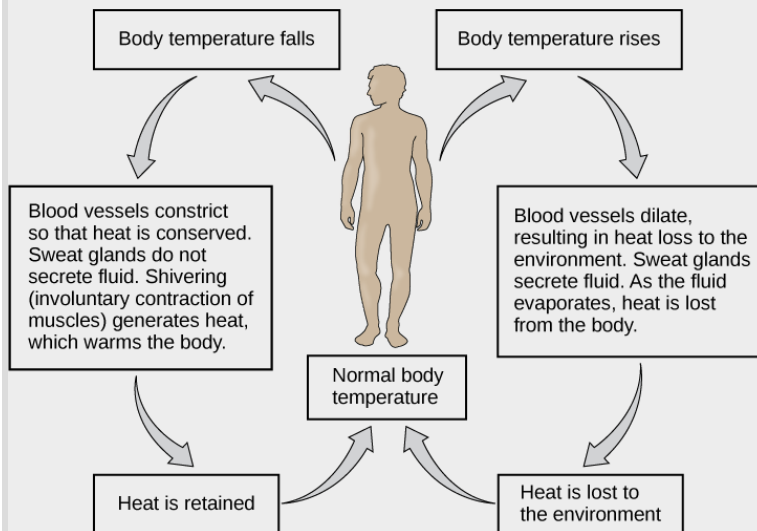
Many animals, especially mammals, use metabolic waste heat as a heat source. When muscles are contracted, most of the energy from the ATP used in muscle actions is wasted energy that translates into heat. Severe cold elicits a shivering reflex that generates heat for the body. Many species also have a type of adipose tissue called brown fat that specializes in generating heat.

## Neural Control of Thermoregulation

The nervous system is important to **thermoregulation**, as illustrated in [\[link\]](#). The processes of homeostasis and temperature control are centered in the hypothalamus of the advanced animal brain.

### Note:

#### Art Connection



The body is able to regulate temperature in response to signals from the nervous system.

When bacteria are destroyed by leuckocytes, pyrogens are released into the blood. Pyrogens reset the body's thermostat to a higher temperature, resulting in fever. How might pyrogens cause the body temperature to rise?

The hypothalamus maintains the set point for body temperature through reflexes that cause vasodilation and sweating when the body is too warm, or vasoconstriction and shivering when the body is too cold. It responds to chemicals from the body. When a bacterium is destroyed by phagocytic leukocytes, chemicals called endogenous pyrogens are released into the blood. These pyrogens circulate to the hypothalamus and reset the thermostat. This allows the body's temperature to increase in what is commonly called a fever. An increase in body temperature causes iron to be conserved, which reduces a nutrient needed by bacteria. An increase in body heat also increases the activity of the animal's enzymes and protective cells while inhibiting the enzymes and activity of the invading microorganisms. Finally, heat itself may also kill the pathogen. A fever that was once thought to be a complication of an infection is now understood to be a normal defense mechanism.

## Section Summary

Homeostasis is a dynamic equilibrium that is maintained in body tissues and organs. It is dynamic because it is constantly adjusting to the changes that the systems encounter. It is in equilibrium because body functions are kept within a normal range, with some fluctuations around a set point for the processes.

## Art Connections

### Exercise:

#### Problem:

[\[link\]](#) State whether each of the following processes are regulated by a positive feedback loop or a negative feedback loop.

- a. A person feels satiated after eating a large meal.
  - b. The blood has plenty of red blood cells. As a result, erythropoietin, a hormone that stimulates the production of new red blood cells, is no longer released from the kidney.
- 

**Solution:**

[\[link\]](#) Both processes are the result of negative feedback loops. Negative feedback loops, which tend to keep a system at equilibrium, are more common than positive feedback loops.

**Exercise:**

**Problem:**

[\[link\]](#) When bacteria are destroyed by leuckocytes, pyrogens are released into the blood. Pyrogens reset the body's thermostat to a higher temperature, resulting in fever. How might pyrogens cause the body temperature to rise?

---

**Solution:**

[\[link\]](#) Pyrogens increase body temperature by causing the blood vessels to constrict, inducing shivering, and stopping sweat glands from secreting fluid.

## Review Questions

**Exercise:**

**Problem:**

When faced with a sudden drop in environmental temperature, an endothermic animal will:

- a. experience a drop in its body temperature
- b. wait to see if it goes lower
- c. increase muscle activity to generate heat

d. add fur or fat to increase insulation

---

**Solution:**

C

**Exercise:**

**Problem:** Which is an example of negative feedback?

- a. lowering of blood glucose after a meal
- b. blood clotting after an injury
- c. lactation during nursing
- d. uterine contractions during labor

---

**Solution:**

A

**Exercise:**

**Problem:**

Which method of heat exchange occurs during direct contact between the source and animal?

- a. radiation
- b. evaporation
- c. convection
- d. conduction

---

**Solution:**

D

**Exercise:**

**Problem:** The body's thermostat is located in the \_\_\_\_\_.

- a. homeostatic receptor
- b. hypothalamus
- c. medulla
- d. vasodilation center

---

**Solution:**

B

## Free Response

**Exercise:**

**Problem:**

Why are negative feedback loops used to control body homeostasis?

---

**Solution:**

An adjustment to a change in the internal or external environment requires a change in the direction of the stimulus. A negative feedback loop accomplishes this, while a positive feedback loop would continue the stimulus and result in harm to the animal.

**Exercise:**

**Problem:** Why is a fever a “good thing” during a bacterial infection?

---

**Solution:**

Mammalian enzymes increase activity to the point of denaturation, increasing the chemical activity of the cells involved. Bacterial enzymes have a specific temperature for their most efficient activity and are inhibited at either higher or lower temperatures. Fever results

in an increase in the destruction of the invading bacteria by increasing the effectiveness of body defenses and an inhibiting bacterial metabolism.

**Exercise:****Problem:**

How is a condition such as diabetes a good example of the failure of a set point in humans?

---

**Solution:**

Diabetes is often associated with a lack in production of insulin. Without insulin, blood glucose levels go up after a meal, but never go back down to normal levels.

**Glossary**

acclimatization

alteration in a body system in response to environmental change

alteration

change of the set point in a homeostatic system

homeostasis

dynamic equilibrium maintaining appropriate body functions

negative feedback loop

feedback to a control mechanism that increases or decreases a stimulus instead of maintaining it

positive feedback loop

feedback to a control mechanism that continues the direction of a stimulus

set point

midpoint or target point in homeostasis



thermoregulation  
regulation of body temperature

## Introduction

class="introduction"

For humans,  
fruits and  
vegetables  
are  
important in  
maintaining  
a balanced  
diet. (credit:  
modificatio  
n of work  
by Julie  
Rybarczyk)



All living organisms need nutrients to survive. While plants can obtain the molecules required for cellular function through the process of photosynthesis, most animals obtain their nutrients by the consumption of other organisms. At the cellular level, the biological molecules necessary for animal function are amino acids, lipid molecules, nucleotides, and simple sugars. However, the food consumed consists of protein, fat, and complex carbohydrates. Animals must convert these macromolecules into the simple molecules required for maintaining cellular functions, such as assembling new molecules, cells, and tissues. The conversion of the food consumed to the nutrients required is a multi-step process involving digestion and absorption. During digestion, food particles are broken down to smaller components, and later, they are absorbed by the body.

One of the challenges in human nutrition is maintaining a balance between food intake, storage, and energy expenditure. Imbalances can have serious health consequences. For example, eating too much food while not expending much energy leads to obesity, which in turn will increase the risk of developing illnesses such as type-2 diabetes and cardiovascular disease. The recent rise in obesity and related diseases makes understanding the role of diet and nutrition in maintaining good health all the more important.

## Digestive Systems

By the end of this section, you will be able to:

- Explain the processes of digestion and absorption
- Compare and contrast different types of digestive systems
- Explain the specialized functions of the organs involved in processing food in the body
- Describe the ways in which organs work together to digest food and absorb nutrients

Animals obtain their nutrition from the consumption of other organisms. Depending on their diet, animals can be classified into the following categories: plant eaters (herbivores), meat eaters (carnivores), and those that eat both plants and animals (omnivores). The nutrients and macromolecules present in food are not immediately accessible to the cells. There are a number of processes that modify food within the animal body in order to make the nutrients and organic molecules accessible for cellular function. As animals evolved in complexity of form and function, their digestive systems have also evolved to accommodate their various dietary needs.

## Herbivores, Omnivores, and Carnivores

**Herbivores** are animals whose primary food source is plant-based. Examples of herbivores, as shown in [\[link\]](#) include vertebrates like deer, koalas, and some bird species, as well as invertebrates such as crickets and caterpillars. These animals have evolved digestive systems capable of handling large amounts of plant material. Herbivores can be further classified into frugivores (fruit-eaters), granivores (seed eaters), nectivores (nectar feeders), and folivores (leaf eaters).



(a)



(b)

Herbivores, like this (a) mule deer and (b) monarch caterpillar, eat primarily plant material. (credit a: modification of work by Bill Ebbesen; credit b: modification of work by Doug Bowman)

**Carnivores** are animals that eat other animals. The word carnivore is derived from Latin and literally means “meat eater.” Wild cats such as lions, shown in [\[link\]](#)a and tigers are examples of vertebrate carnivores, as are snakes and sharks, while invertebrate carnivores include sea stars, spiders, and ladybugs, shown in [\[link\]](#)b. Obligate carnivores are those that rely entirely on animal flesh to obtain their nutrients; examples of obligate carnivores are members of the cat family, such as lions and cheetahs. Facultative carnivores are those that also eat non-animal food in addition to animal food. Note that there is no clear line that differentiates facultative carnivores from omnivores; dogs would be considered facultative carnivores.





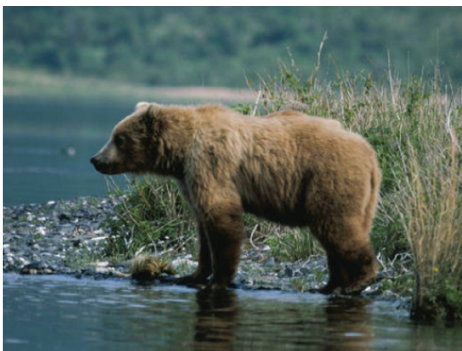
(a)



(b)

Carnivores like the (a) lion eat primarily meat. The (b) ladybug is also a carnivore that consumes small insects called aphids. (credit a: modification of work by Kevin Pluck; credit b: modification of work by Jon Sullivan)

**Omnivores** are animals that eat both plant- and animal-derived food. In Latin, omnivore means to eat everything. Humans, bears (shown in [\[link\]](#)a), and chickens are example of vertebrate omnivores; invertebrate omnivores include cockroaches and crayfish (shown in [\[link\]](#)b).



(a)



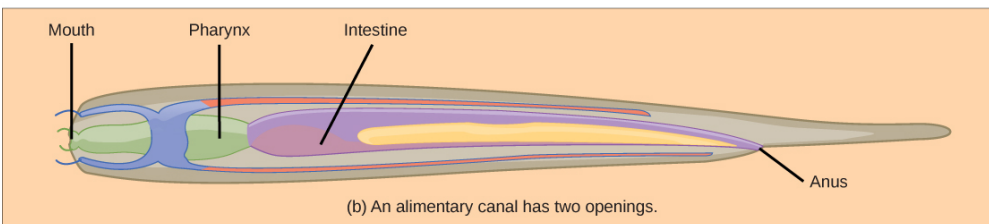
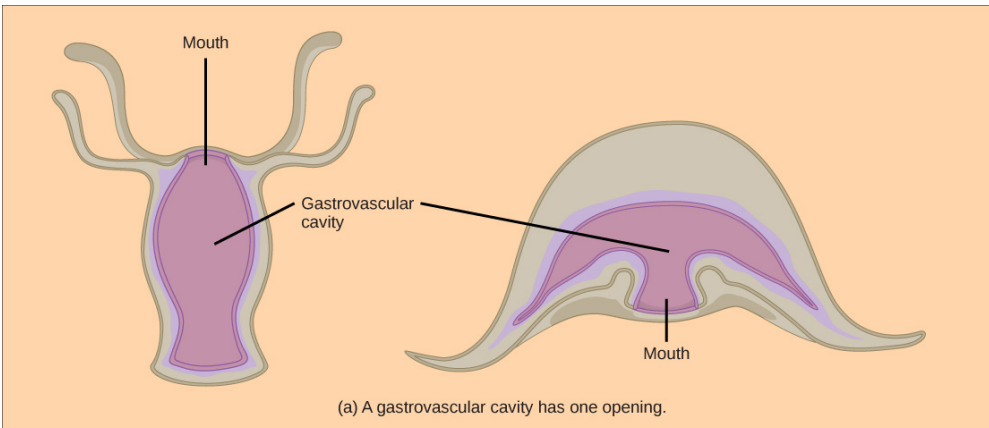
(b)

Omnivores like the (a) bear and (b) crayfish eat both plant and animal based food. (credit a: modification of work by Dave Menke; credit b: modification of work by Jon Sullivan)

## Invertebrate Digestive Systems

Animals have evolved different types of digestive systems to aid in the digestion of the different foods they consume. The simplest example is that of a **gastrovascular cavity** and is found in organisms with only one opening for digestion. Platyhelminthes (flatworms), Ctenophora (comb jellies), and Cnidaria (coral, jelly fish, and sea anemones) use this type of digestion. Gastrovascular cavities, as shown in [\[link\]a](#), are typically a blind tube or cavity with only one opening, the “mouth”, which also serves as an “anus”. Ingested material enters the mouth and passes through a hollow, tubular cavity. Cells within the cavity secrete digestive enzymes that break down the food. The food particles are engulfed by the cells lining the gastrovascular cavity.

The **alimentary canal**, shown in [\[link\]b](#), is a more advanced system: it consists of one tube with a mouth at one end and an anus at the other. Earthworms are an example of an animal with an alimentary canal. Once the food is ingested through the mouth, it passes through the esophagus and is stored in an organ called the crop; then it passes into the gizzard where it is churned and digested. From the gizzard, the food passes through the intestine, the nutrients are absorbed, and the waste is eliminated as feces, called castings, through the anus.



(a) A gastrovascular cavity has a single opening through which food is ingested and waste is excreted, as shown in this hydra and in this jellyfish medusa. (b) An alimentary canal has two openings: a mouth for ingesting food, and an anus for eliminating waste, as shown in this nematode.

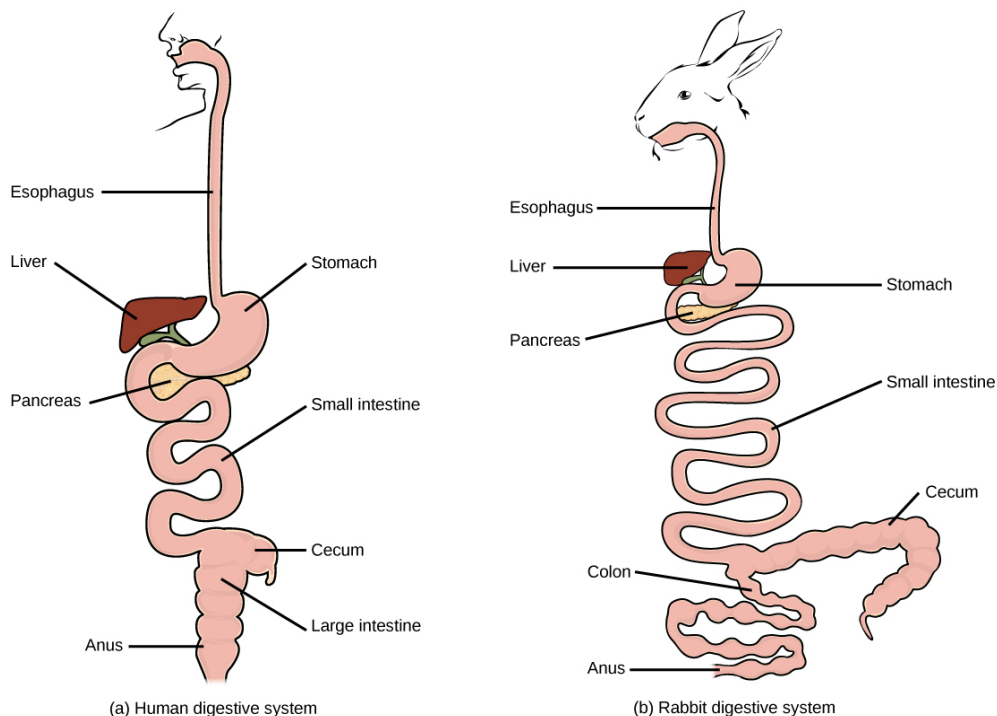
## Vertebrate Digestive Systems

Vertebrates have evolved more complex digestive systems to adapt to their dietary needs. Some animals have a single stomach, while others have multi-chambered stomachs. Birds have developed a digestive system adapted to eating unmasticated food.

### Monogastric: Single-chambered Stomach



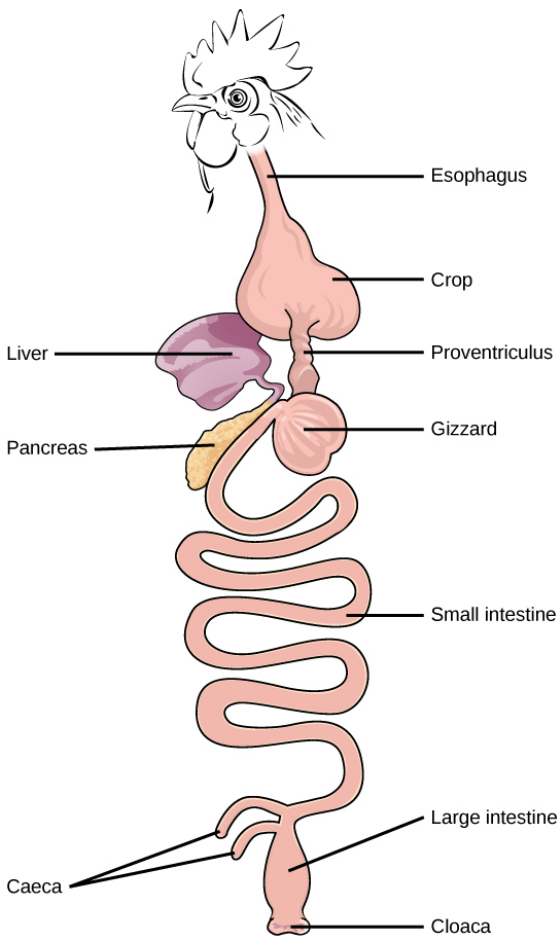
As the word **monogastric** suggests, this type of digestive system consists of one (“mono”) stomach chamber (“gastric”). Humans and many animals have a monogastric digestive system as illustrated in [\[link\]](#)ab. The process of digestion begins with the mouth and the intake of food. The teeth play an important role in masticating (chewing) or physically breaking down food into smaller particles. The enzymes present in saliva also begin to chemically break down food. The esophagus is a long tube that connects the mouth to the stomach. Using peristalsis, or wave-like smooth muscle contractions, the muscles of the esophagus push the food towards the stomach. In order to speed up the actions of enzymes in the stomach, the stomach is an extremely acidic environment, with a pH between 1.5 and 2.5. The gastric juices, which include enzymes in the stomach, act on the food particles and continue the process of digestion. Further breakdown of food takes place in the small intestine where enzymes produced by the liver, the small intestine, and the pancreas continue the process of digestion. The nutrients are absorbed into the blood stream across the epithelial cells lining the walls of the small intestines. The waste material travels on to the large intestine where water is absorbed and the drier waste material is compacted into feces; it is stored until it is excreted through the rectum.



(a) Humans and herbivores, such as the (b) rabbit, have a monogastric digestive system. However, in the rabbit the small intestine and cecum are enlarged to allow more time to digest plant material. The enlarged organ provides more surface area for absorption of nutrients. Rabbits digest their food twice: the first time food passes through the digestive system, it collects in the cecum, and then it passes as soft feces called cecotrophes. The rabbit re-ingests these cecotrophes to further digest them.

## Avian

Birds face special challenges when it comes to obtaining nutrition from food. They do not have teeth and so their digestive system, shown in [\[link\]](#), must be able to process un-masticated food. Birds have evolved a variety of beak types that reflect the vast variety in their diet, ranging from seeds and insects to fruits and nuts. Because most birds fly, their metabolic rates are high in order to efficiently process food and keep their body weight low. The stomach of birds has two chambers: the **proventriculus**, where gastric juices are produced to digest the food before it enters the stomach, and the **gizzard**, where the food is stored, soaked, and mechanically ground. The undigested material forms food pellets that are sometimes regurgitated. Most of the chemical digestion and absorption happens in the intestine and the waste is excreted through the cloaca.



The avian esophagus has a pouch, called a crop, which stores food. Food passes from the crop to the first of two stomachs, called the proventriculus, which contains digestive juices that break down food. From the proventriculus, the food enters the second stomach, called the gizzard, which grinds food. Some birds swallow stones or grit, which are stored in the gizzard, to aid the grinding process. Birds do not have separate

openings to excrete urine and feces. Instead, uric acid from the kidneys is secreted into the large intestine and combined with waste from the digestive process. This waste is excreted through an opening called the cloaca.

**Note:**

**Evolution Connection**

**Avian Adaptations**

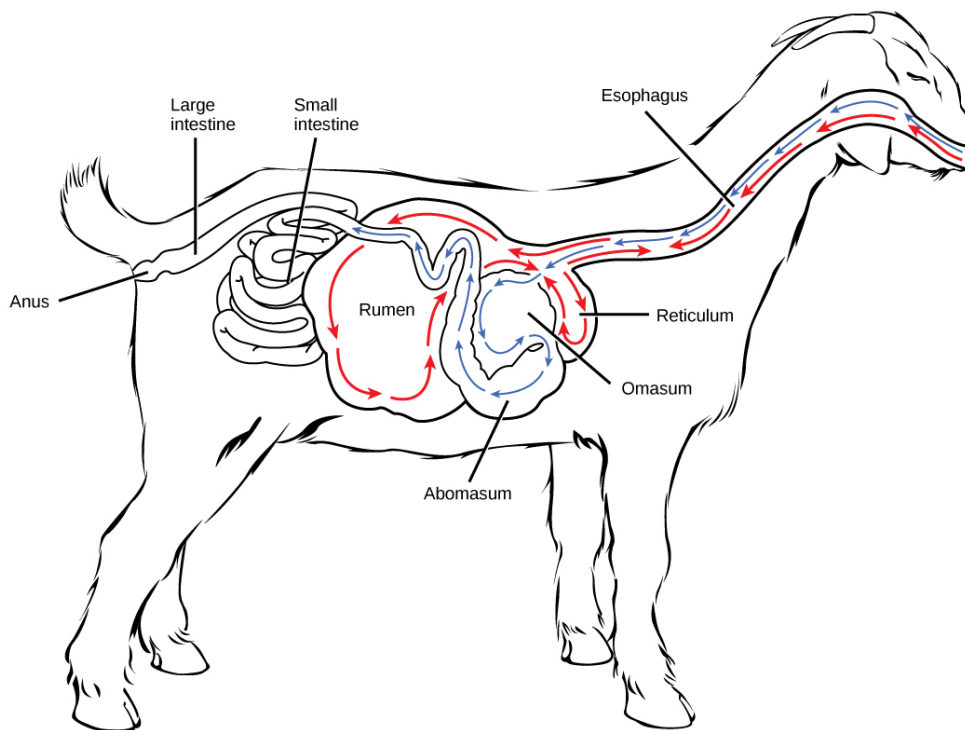
Birds have a highly efficient, simplified digestive system. Recent fossil evidence has shown that the evolutionary divergence of birds from other land animals was characterized by streamlining and simplifying the digestive system. Unlike many other animals, birds do not have teeth to chew their food. In place of lips, they have sharp pointy beaks. The horny beak, lack of jaws, and the smaller tongue of the birds can be traced back to their dinosaur ancestors. The emergence of these changes seems to coincide with the inclusion of seeds in the bird diet. Seed-eating birds have beaks that are shaped for grabbing seeds and the two-compartment stomach allows for delegation of tasks. Since birds need to remain light in order to fly, their metabolic rates are very high, which means they digest their food very quickly and need to eat often. Contrast this with the ruminants, where the digestion of plant matter takes a very long time.

**Ruminants**

**Ruminants** are mainly herbivores like cows, sheep, and goats, whose entire diet consists of eating large amounts of **roughage** or fiber. They have evolved digestive systems that help them digest vast amounts of cellulose. An interesting feature of the ruminants' mouth is that they do not have

upper incisor teeth. They use their lower teeth, tongue and lips to tear and chew their food. From the mouth, the food travels to the esophagus and on to the stomach.

To help digest the large amount of plant material, the stomach of the ruminants is a multi-chambered organ, as illustrated in [\[link\]](#). The four compartments of the stomach are called the rumen, reticulum, omasum, and abomasum. These chambers contain many microbes that break down cellulose and ferment ingested food. The abomasum is the “true” stomach and is the equivalent of the monogastric stomach chamber where gastric juices are secreted. The four-compartment gastric chamber provides larger space and the microbial support necessary to digest plant material in ruminants. The fermentation process produces large amounts of gas in the stomach chamber, which must be eliminated. As in other animals, the small intestine plays an important role in nutrient absorption, and the large intestine helps in the elimination of waste.



Ruminant animals, such as goats and cows, have four stomachs. The first two stomachs, the rumen and the

reticulum, contain prokaryotes and protists that are able to digest cellulose fiber. The ruminant regurgitates cud from the reticulum, chews it, and swallows it into a third stomach, the omasum, which removes water. The cud then passes onto the fourth stomach, the abomasum, where it is digested by enzymes produced by the ruminant.

### **Pseudo-ruminants**

Some animals, such as camels and alpacas, are pseudo-ruminants. They eat a lot of plant material and roughage. Digesting plant material is not easy because plant cell walls contain the polymeric sugar molecule cellulose. The digestive enzymes of these animals cannot break down cellulose, but microorganisms present in the digestive system can. Therefore, the digestive system must be able to handle large amounts of roughage and break down the cellulose. Pseudo-ruminants have a three-chamber stomach in the digestive system. However, their cecum—a pouched organ at the beginning of the large intestine containing many microorganisms that are necessary for the digestion of plant materials—is large and is the site where the roughage is fermented and digested. These animals do not have a rumen but have an omasum, abomasum, and reticulum.

## **Parts of the Digestive System**

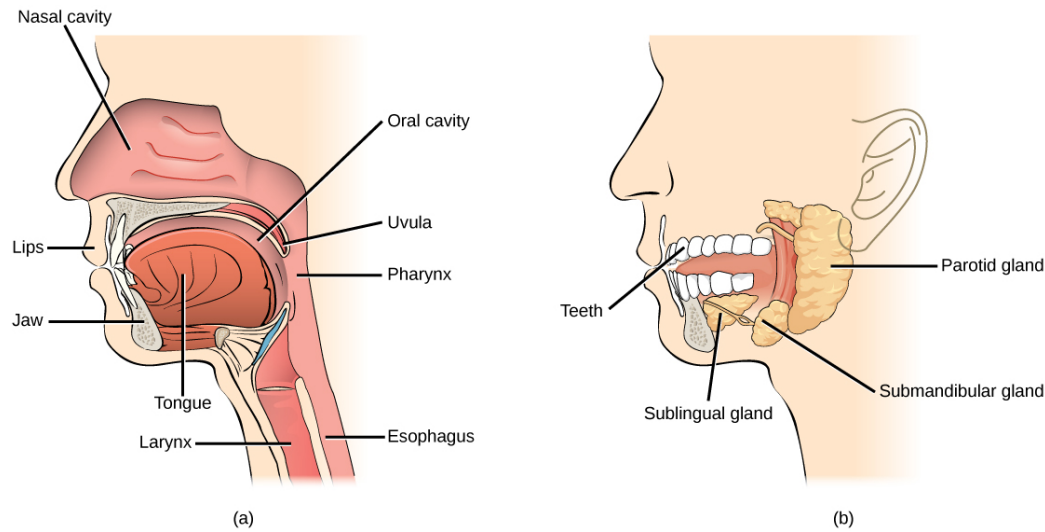
The vertebrate digestive system is designed to facilitate the transformation of food matter into the nutrient components that sustain organisms.

### **Oral Cavity**

The oral cavity, or mouth, is the point of entry of food into the digestive system, illustrated in [\[link\]](#). The food consumed is broken into smaller

particles by mastication, the chewing action of the teeth. All mammals have teeth and can chew their food.

The extensive chemical process of digestion begins in the mouth. As food is being chewed, saliva, produced by the salivary glands, mixes with the food. Saliva is a watery substance produced in the mouths of many animals. There are three major glands that secrete saliva—the parotid, the submandibular, and the sublingual. Saliva contains mucus that moistens food and buffers the pH of the food. Saliva also contains immunoglobulins and lysozymes, which have antibacterial action to reduce tooth decay by inhibiting growth of some bacteria. Saliva also contains an enzyme called **salivary amylase** that begins the process of converting starches in the food into a disaccharide called maltose. Another enzyme called **lipase** is produced by the cells in the tongue. Lipases are a class of enzymes that can break down triglycerides. The lingual lipase begins the breakdown of fat components in the food. The chewing and wetting action provided by the teeth and saliva prepare the food into a mass called the **bolus** for swallowing. The tongue helps in swallowing—moving the bolus from the mouth into the pharynx. The pharynx opens to two passageways: the trachea, which leads to the lungs, and the esophagus, which leads to the stomach. The trachea has an opening called the glottis, which is covered by a cartilaginous flap called the epiglottis. When swallowing, the epiglottis closes the glottis and food passes into the esophagus and not the trachea. This arrangement allows food to be kept out of the trachea.



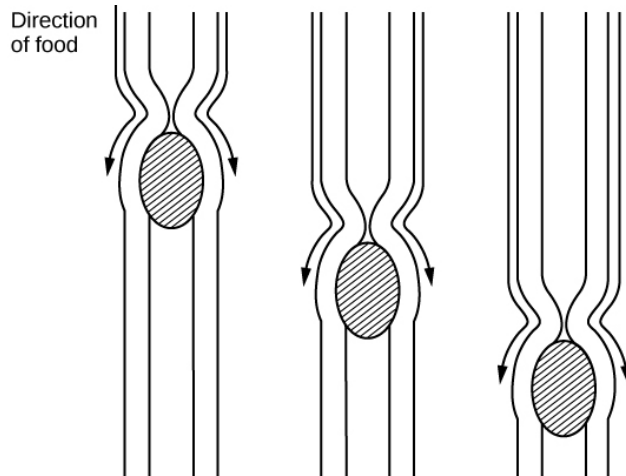
Digestion of food begins in the (a) oral cavity. Food is masticated by teeth and moistened by saliva secreted from the (b) salivary glands. Enzymes in the saliva begin to digest starches and fats. With the help of the tongue, the resulting bolus is moved into the esophagus by swallowing.

(credit: modification of work by the National Cancer Institute)

## Esophagus

The **esophagus** is a tubular organ that connects the mouth to the stomach. The chewed and softened food passes through the esophagus after being swallowed. The smooth muscles of the esophagus undergo a series of wave like movements called **peristalsis** that push the food toward the stomach, as illustrated in [\[link\]](#). The peristalsis wave is unidirectional—it moves food from the mouth to the stomach, and reverse movement is not possible. The peristaltic movement of the esophagus is an involuntary reflex; it takes place in response to the act of swallowing.





The esophagus transfers food from the mouth to the stomach through peristaltic movements.

A ring-like muscle called a **sphincter** forms valves in the digestive system. The gastro-esophageal sphincter is located at the stomach end of the esophagus. In response to swallowing and the pressure exerted by the bolus of food, this sphincter opens, and the bolus enters the stomach. When there is no swallowing action, this sphincter is shut and prevents the contents of the stomach from traveling up the esophagus. Many animals have a true sphincter; however, in humans, there is no true sphincter, but the esophagus remains closed when there is no swallowing action. Acid reflux or “heartburn” occurs when the acidic digestive juices escape into the esophagus.

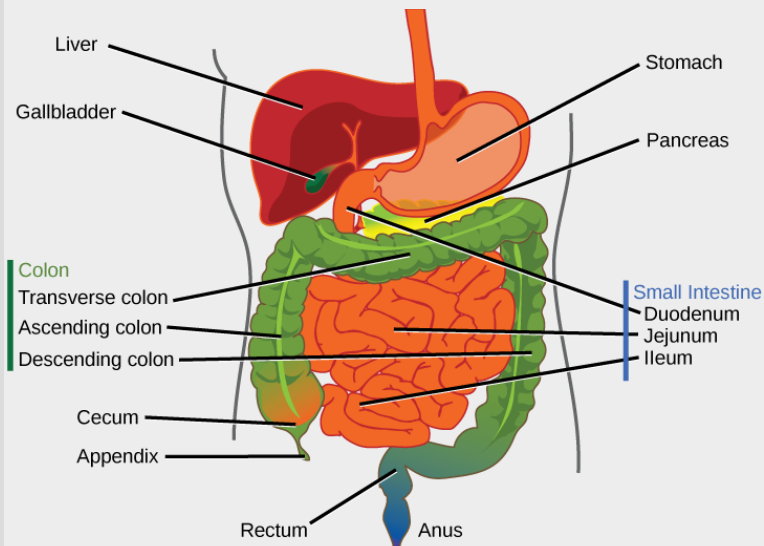
## **Stomach**

A large part of digestion occurs in the stomach, shown in [\[link\]](#). The **stomach** is a saclike organ that secretes gastric digestive juices. The pH in the stomach is between 1.5 and 2.5. This highly acidic environment is required for the chemical breakdown of food and the extraction of nutrients. When empty, the stomach is a rather small organ; however, it can expand to

up to 20 times its resting size when filled with food. This characteristic is particularly useful for animals that need to eat when food is available.

**Note:**

**Art Connection**



The human stomach has an extremely acidic environment where most of the protein gets digested. (credit: modification of work by Mariana Ruiz Villareal)

Which of the following statements about the digestive system is false?

- a. Chyme is a mixture of food and digestive juices that is produced in the stomach.
- b. Food enters the large intestine before the small intestine.
- c. In the small intestine, chyme mixes with bile, which emulsifies fats.
- d. The stomach is separated from the small intestine by the pyloric sphincter.

The stomach is also the major site for protein digestion in animals other than ruminants. Protein digestion is mediated by an enzyme called pepsin in the stomach chamber. **Pepsin** is secreted by the chief cells in the stomach in an inactive form called **pepsinogen**. Pepsin breaks peptide bonds and cleaves proteins into smaller polypeptides; it also helps activate more pepsinogen, starting a positive feedback mechanism that generates more pepsin. Another cell type—parietal cells—secrete hydrogen and chloride ions, which combine in the lumen to form hydrochloric acid, the primary acidic component of the stomach juices. Hydrochloric acid helps to convert the inactive pepsinogen to pepsin. The highly acidic environment also kills many microorganisms in the food and, combined with the action of the enzyme pepsin, results in the hydrolysis of protein in the food. Chemical digestion is facilitated by the churning action of the stomach. Contraction and relaxation of smooth muscles mixes the stomach contents about every 20 minutes. The partially digested food and gastric juice mixture is called **chyme**. Chyme passes from the stomach to the small intestine. Further protein digestion takes place in the small intestine. Gastric emptying occurs within two to six hours after a meal. Only a small amount of chyme is released into the small intestine at a time. The movement of chyme from the stomach into the small intestine is regulated by the pyloric sphincter.

When digesting protein and some fats, the stomach lining must be protected from getting digested by pepsin. There are two points to consider when describing how the stomach lining is protected. First, as previously mentioned, the enzyme pepsin is synthesized in the inactive form. This protects the chief cells, because pepsinogen does not have the same enzyme functionality of pepsin. Second, the stomach has a thick mucus lining that protects the underlying tissue from the action of the digestive juices. When this mucus lining is ruptured, ulcers can form in the stomach. Ulcers are open wounds in or on an organ caused by bacteria (*Helicobacter pylori*) when the mucus lining is ruptured and fails to reform.

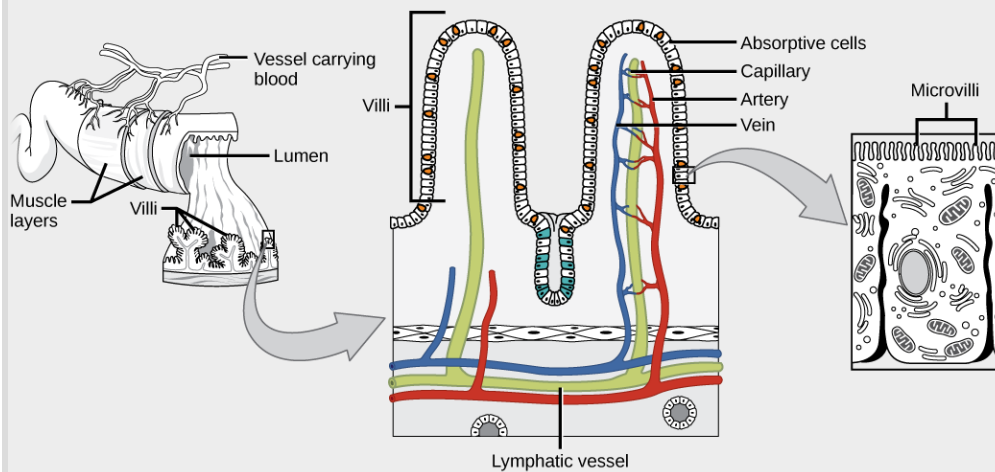
## **Small Intestine**

Chyme moves from the stomach to the small intestine. The **small intestine** is the organ where the digestion of protein, fats, and carbohydrates is

completed. The small intestine is a long tube-like organ with a highly folded surface containing finger-like projections called the **villi**. The apical surface of each villus has many microscopic projections called microvilli. These structures, illustrated in [\[link\]](#), are lined with epithelial cells on the luminal side and allow for the nutrients to be absorbed from the digested food and absorbed into the blood stream on the other side. The villi and microvilli, with their many folds, increase the surface area of the intestine and increase absorption efficiency of the nutrients. Absorbed nutrients in the blood are carried into the hepatic portal vein, which leads to the liver. There, the liver regulates the distribution of nutrients to the rest of the body and removes toxic substances, including drugs, alcohol, and some pathogens.

**Note:**

**Art Connection**



Villi are folds on the small intestine lining that increase the surface area to facilitate the absorption of nutrients.

Which of the following statements about the small intestine is false?

- Absorptive cells that line the small intestine have microvilli, small projections that increase surface area and aid in the absorption of food.

- b. The inside of the small intestine has many folds, called villi.
- c. Microvilli are lined with blood vessels as well as lymphatic vessels.
- d. The inside of the small intestine is called the lumen.

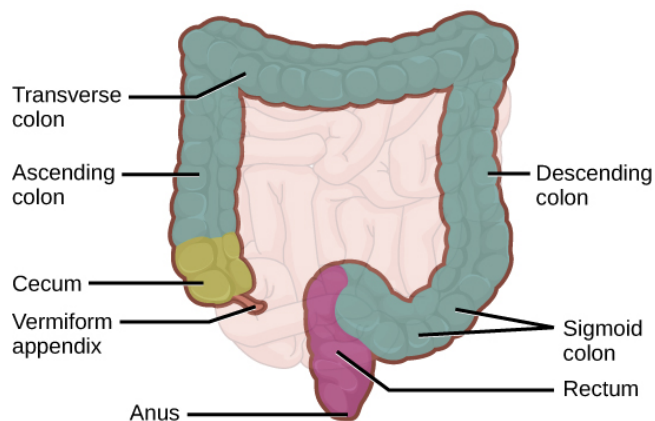
The human small intestine is over 6m long and is divided into three parts: the duodenum, the jejunum, and the ileum. The “C-shaped,” fixed part of the small intestine is called the **duodenum** and is shown in [\[link\]](#). The duodenum is separated from the stomach by the pyloric sphincter which opens to allow chyme to move from the stomach to the duodenum. In the duodenum, chyme is mixed with pancreatic juices in an alkaline solution rich in bicarbonate that neutralizes the acidity of chyme and acts as a buffer. Pancreatic juices also contain several digestive enzymes. Digestive juices from the pancreas, liver, and gallbladder, as well as from gland cells of the intestinal wall itself, enter the duodenum. **Bile** is produced in the liver and stored and concentrated in the gallbladder. Bile contains bile salts which emulsify lipids while the pancreas produces enzymes that catabolize starches, disaccharides, proteins, and fats. These digestive juices break down the food particles in the chyme into glucose, triglycerides, and amino acids. Some chemical digestion of food takes place in the duodenum. Absorption of fatty acids also takes place in the duodenum.

The second part of the small intestine is called the **jejunum**, shown in [\[link\]](#). Here, hydrolysis of nutrients is continued while most of the carbohydrates and amino acids are absorbed through the intestinal lining. The bulk of chemical digestion and nutrient absorption occurs in the jejunum.

The **ileum**, also illustrated in [\[link\]](#) is the last part of the small intestine and here the bile salts and vitamins are absorbed into blood stream. The undigested food is sent to the colon from the ileum via peristaltic movements of the muscle. The ileum ends and the large intestine begins at the ileocecal valve. The vermiform, “worm-like,” appendix is located at the ileocecal valve. The appendix of humans secretes no enzymes and has an insignificant role in immunity.

## Large Intestine

The **large intestine**, illustrated in [\[link\]](#), reabsorbs the water from the undigested food material and processes the waste material. The human large intestine is much smaller in length compared to the small intestine but larger in diameter. It has three parts: the cecum, the colon, and the rectum. The cecum joins the ileum to the colon and is the receiving pouch for the waste matter. The colon is home to many bacteria or “intestinal flora” that aid in the digestive processes. The colon can be divided into four regions, the ascending colon, the transverse colon, the descending colon and the sigmoid colon. The main functions of the colon are to extract the water and mineral salts from undigested food, and to store waste material. Carnivorous mammals have a shorter large intestine compared to herbivorous mammals due to their diet.



The large intestine reabsorbs water from undigested food and stores waste material until it is eliminated.

## Rectum and Anus

The **rectum** is the terminal end of the large intestine, as shown in [\[link\]](#). The primary role of the rectum is to store the feces until defecation. The feces are propelled using peristaltic movements during elimination. The **anus** is an opening at the far-end of the digestive tract and is the exit point for the waste material. Two sphincters between the rectum and anus control elimination: the inner sphincter is involuntary and the outer sphincter is voluntary.

## Accessory Organs

The organs discussed above are the organs of the digestive tract through which food passes. Accessory organs are organs that add secretions (enzymes) that catabolize food into nutrients. Accessory organs include salivary glands, the liver, the pancreas, and the gallbladder. The liver, pancreas, and gallbladder are regulated by hormones in response to the food consumed.

The **liver** is the largest internal organ in humans and it plays a very important role in digestion of fats and detoxifying blood. The liver produces bile, a digestive juice that is required for the breakdown of fatty components of the food in the duodenum. The liver also processes the vitamins and fats and synthesizes many plasma proteins.

The **pancreas** is another important gland that secretes digestive juices. The chyme produced from the stomach is highly acidic in nature; the pancreatic juices contain high levels of bicarbonate, an alkali that neutralizes the acidic chyme. Additionally, the pancreatic juices contain a large variety of enzymes that are required for the digestion of protein and carbohydrates.

The **gallbladder** is a small organ that aids the liver by storing bile and concentrating bile salts. When chyme containing fatty acids enters the duodenum, the bile is secreted from the gallbladder into the duodenum.

## Section Summary

Different animals have evolved different types of digestive systems specialized to meet their dietary needs. Humans and many other animals have monogastric digestive systems with a single-chambered stomach. Birds have evolved a digestive system that includes a gizzard where the food is crushed into smaller pieces. This compensates for their inability to masticate. Ruminants that consume large amounts of plant material have a multi-chambered stomach that digests roughage. Pseudo-ruminants have similar digestive processes as ruminants but do not have the four-compartment stomach. Processing food involves ingestion (eating), digestion (mechanical and enzymatic breakdown of large molecules), absorption (cellular uptake of nutrients), and elimination (removal of undigested waste as feces).

Many organs work together to digest food and absorb nutrients. The mouth is the point of ingestion and the location where both mechanical and chemical breakdown of food begins. Saliva contains an enzyme called amylase that breaks down carbohydrates. The food bolus travels through the esophagus by peristaltic movements to the stomach. The stomach has an extremely acidic environment. An enzyme called pepsin digests protein in the stomach. Further digestion and absorption take place in the small intestine. The large intestine reabsorbs water from the undigested food and stores waste until elimination.

## Art Connections

### Exercise:

#### Problem:

[\[link\]](#) Which of the following statements about the digestive system is false?

- a. Chyme is a mixture of food and digestive juices that is produced in the stomach.
- b. Food enters the large intestine before the small intestine.
- c. In the small intestine, chyme mixes with bile, which emulsifies fats.



- d. The stomach is separated from the small intestine by the pyloric sphincter.
- 

**Solution:**

[\[link\]](#) B

**Exercise:**

**Problem:**

[\[link\]](#) Which of the following statements about the small intestine is false?

- a. Absorptive cells that line the small intestine have microvilli, small projections that increase surface area and aid in the absorption of food.
  - b. The inside of the small intestine has many folds, called villi.
  - c. Microvilli are lined with blood vessels as well as lymphatic vessels.
  - d. The inside of the small intestine is called the lumen.
- 

**Solution:**

[\[link\]](#) C

## Review Questions

**Exercise:**

**Problem:** Which of the following is a pseudo-ruminant?

- a. cow
- b. pig
- c. crow
- d. horse

---

**Solution:**

D

**Exercise:**

**Problem:** Which of the following statements is untrue?

- a. Roughage takes a long time to digest.
- b. Birds eat large quantities at one time so that they can fly long distances.
- c. Cows do not have upper teeth.
- d. In pseudo-ruminants, roughage is digested in the cecum.

---

**Solution:**

B

**Exercise:**

**Problem:** The acidic nature of chyme is neutralized by \_\_\_\_\_.

- a. potassium hydroxide
- b. sodium hydroxide
- c. bicarbonates
- d. vinegar

---

**Solution:**

C

**Exercise:**

**Problem:**

The digestive juices from the liver are delivered to the \_\_\_\_\_.

- a. stomach
- b. liver
- c. duodenum
- d. colon

---

**Solution:**

C

## Free Response

**Exercise:**

**Problem:**

How does the polygastric digestive system aid in digesting roughage?

---

**Solution:**

Animals with a polygastric digestive system have a multi-chambered stomach. The four compartments of the stomach are called the rumen, reticulum, omasum, and abomasum. These chambers contain many microbes that break down the cellulose and ferment the ingested food. The abomasum is the “true” stomach and is the equivalent of a monogastric stomach chamber where gastric juices are secreted. The four-compartment gastric chamber provides larger space and the microbial support necessary for ruminants to digest plant material.

**Exercise:**

**Problem:**How do birds digest their food in the absence of teeth?

---

**Solution:**

Birds have a stomach chamber called a gizzard. Here, the food is stored, soaked, and ground into finer particles, often using pebbles.

Once this process is complete, the digestive juices take over in the proventriculus and continue the digestive process.

**Exercise:**

**Problem:**What is the role of the accessory organs in digestion?

---

**Solution:**

Accessory organs play an important role in producing and delivering digestive juices to the intestine during digestion and absorption. Specifically, the salivary glands, liver, pancreas, and gallbladder play important roles. Malfunction of any of these organs can lead to disease states.

**Exercise:**

**Problem:**Explain how the villi and microvilli aid in absorption.

---

**Solution:**

The villi and microvilli are folds on the surface of the small intestine. These folds increase the surface area of the intestine and provide more area for the absorption of nutrients.

## Glossary

alimentary canal

tubular digestive system with a mouth and anus

anus

exit point for waste material

bile

digestive juice produced by the liver; important for digestion of lipids

bolus

mass of food resulting from chewing action and wetting by saliva

carnivore

animal that consumes animal flesh

chyme

mixture of partially digested food and stomach juices

duodenum

first part of the small intestine where a large part of digestion of carbohydrates and fats occurs

esophagus

tubular organ that connects the mouth to the stomach

gallbladder

organ that stores and concentrates bile

gastrovascular cavity

digestive system consisting of a single opening

gizzard

muscular organ that grinds food

herbivore

animal that consumes strictly plant diet

ileum

last part of the small intestine; connects the small intestine to the large intestine; important for absorption of B-12

jejunum

second part of the small intestine

large intestine

digestive system organ that reabsorbs water from undigested material and processes waste matter

lipase

enzyme that chemically breaks down lipids

liver

organ that produces bile for digestion and processes vitamins and lipids

monogastric

digestive system that consists of a single-chambered stomach

omnivore

animal that consumes both plants and animals

pancreas

gland that secretes digestive juices

pepsin

enzyme found in the stomach whose main role is protein digestion

pepsinogen

inactive form of pepsin

peristalsis

wave-like movements of muscle tissue

proventriculus

glandular part of a bird's stomach

rectum

area of the body where feces is stored until elimination

roughage

component of food that is low in energy and high in fiber

ruminant

animal with a stomach divided into four compartments

salivary amylase

enzyme found in saliva, which converts carbohydrates to maltose

small intestine

organ where digestion of protein, fats, and carbohydrates is completed

sphincter

band of muscle that controls movement of materials throughout the digestive tract

stomach

saclike organ containing acidic digestive juices

villi

foldings on the inner surface of the small intestine whose role is to increase absorption area

## Nutrition and Energy Production

By the end of this section, you will be able to:

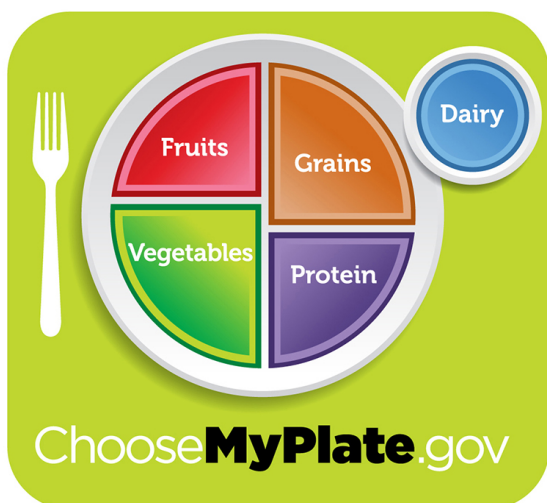
- Explain why an animal's diet should be balanced and meet the needs of the body
- Define the primary components of food
- Describe the essential nutrients required for cellular function that cannot be synthesized by the animal body
- Explain how energy is produced through diet and digestion
- Describe how excess carbohydrates and energy are stored in the body

Given the diversity of animal life on our planet, it is not surprising that the animal diet would also vary substantially. The animal diet is the source of materials needed for building DNA and other complex molecules needed for growth, maintenance, and reproduction; collectively these processes are called biosynthesis. The diet is also the source of materials for ATP production in the cells. The diet must be balanced to provide the minerals and vitamins that are required for cellular function.

## Food Requirements

What are the fundamental requirements of the animal diet? The animal diet should be well balanced and provide nutrients required for bodily function and the minerals and vitamins required for maintaining structure and regulation necessary for good health and reproductive capability. These requirements for a human are illustrated graphically in [\[link\]](#)





For humans, a balanced diet includes fruits, vegetables, grains, and protein. (credit: USDA)

**Note:**

Link to Learning



The first step in ensuring that you are meeting the food requirements of your body is an awareness of the food groups and the nutrients they provide. To learn more about each food group and the recommended daily amounts, explore this [interactive site](#) by the United States Department of Agriculture.

**Note:****Everyday Connection****Let's Move! Campaign**

Obesity is a growing epidemic and the rate of obesity among children is rapidly rising in the United States. To combat childhood obesity and ensure that children get a healthy start in life, first lady Michelle Obama has launched the Let's Move! campaign. The goal of this campaign is to educate parents and caregivers on providing healthy nutrition and encouraging active lifestyles to future generations. This program aims to involve the entire community, including parents, teachers, and healthcare providers to ensure that children have access to healthy foods—more fruits, vegetables, and whole grains—and consume fewer calories from processed foods. Another goal is to ensure that children get physical activity. With the increase in television viewing and stationary pursuits such as video games, sedentary lifestyles have become the norm. Learn more at [www.letsmove.gov](http://www.letsmove.gov).

**Organic Precursors**

The organic molecules required for building cellular material and tissues must come from food. Carbohydrates or sugars are the primary source of organic carbons in the animal body. During digestion, digestible carbohydrates are ultimately broken down into glucose and used to provide energy through metabolic pathways. Complex carbohydrates, including polysaccharides, can be broken down into glucose through biochemical modification; however, humans do not produce the enzyme cellulase and lack the ability to derive glucose from the polysaccharide cellulose. In humans, these molecules provide the fiber required for moving waste through the large intestine and a healthy colon. The intestinal flora in the human gut are able to extract some nutrition from these plant fibers. The excess sugars in the body are converted into glycogen and stored in the liver and muscles for later use. Glycogen stores are used to fuel prolonged exertions, such as long-distance running, and to provide energy during food shortage. Excess glycogen can be converted to fats, which are stored in the lower layer of the skin of mammals for insulation and energy storage. Excess digestible carbohydrates are stored by mammals in order to survive famine and aid in mobility.

Another important requirement is that of nitrogen. Protein catabolism provides a source of organic nitrogen. Amino acids are the building blocks of proteins and protein breakdown provides amino acids that are used for cellular function. The carbon and nitrogen derived from these become the building block for nucleotides, nucleic acids, proteins, cells, and tissues. Excess nitrogen must be excreted as it is toxic. Fats add flavor to food and promote a sense of satiety or fullness. Fatty foods are also significant sources of energy because one gram of fat contains nine calories. Fats are required in the diet to aid the absorption of fat-soluble vitamins and the production of fat-soluble hormones.

## Essential Nutrients

While the animal body can synthesize many of the molecules required for function from the organic precursors, there are some nutrients that need to be consumed from food. These nutrients are termed **essential nutrients**, meaning they must be eaten, and the body cannot produce them.

The omega-3 alpha-linolenic acid and the omega-6 linoleic acid are essential fatty acids needed to make some membrane phospholipids. **Vitamins** are another class of essential organic molecules that are required in small quantities for many enzymes to function and, for this reason, are considered to be co-enzymes. Absence or low levels of vitamins can have a dramatic effect on health, as outlined in [\[link\]](#) and [\[link\]](#). Both fat-soluble and water-soluble vitamins must be obtained from food. **Minerals**, listed in [\[link\]](#), are inorganic essential nutrients that must be obtained from food. Among their many functions, minerals help in structure and regulation and are considered co-factors. Certain amino acids also must be procured from food and cannot be synthesized by the body. These amino acids are the “essential” amino acids. The human body can synthesize only 11 of the 20 required amino acids; the rest must be obtained from food. The essential amino acids are listed in [\[link\]](#).

<b>Water-soluble Essential Vitamins</b>			
<b>Vitamin</b>	<b>Function</b>	<b>Deficiencies Can Lead To</b>	<b>Sources</b>
Vitamin B <sub>1</sub> (Thiamine)	Needed by the body to process lipids, proteins, and carbohydrates Coenzyme removes CO <sub>2</sub> from organic compounds	Muscle weakness, Beriberi: reduced heart function, CNS problems	Milk, meat, dried beans, whole grains
Vitamin B <sub>2</sub> (Riboflavin)	Takes an active role in metabolism, aiding in the conversion of food to energy (FAD and FMN)	Cracks or sores on the outer surface of the lips (cheliosis); inflammation and redness of the tongue; moist, scaly skin inflammation (seborrheic dermatitis)	Meat, eggs, enriched grains, vegetables

Water-soluble Essential Vitamins			
Vitamin	Function	Deficiencies Can Lead To	Sources
Vitamin B <sub>3</sub> (Niacin)	Used by the body to release energy from carbohydrates and to process alcohol; required for the synthesis of sex hormones; component of coenzyme NAD <sup>+</sup> and NADP <sup>+</sup>	Pellagra, which can result in dermatitis, diarrhea, dementia, and death	Meat, eggs, grains, nuts, potatoes
Vitamin B <sub>5</sub> (Pantothenic acid)	Assists in producing energy from foods (lipids, in particular); component of coenzyme A	Fatigue, poor coordination, retarded growth, numbness, tingling of hands and feet	Meat, whole grains, milk, fruits, vegetables
Vitamin B <sub>6</sub> (Pyridoxine)	The principal vitamin for processing amino acids and lipids; also helps convert nutrients into energy	Irritability, depression, confusion, mouth sores or ulcers, anemia, muscular twitching	Meat, dairy products, whole grains, orange juice

<b>Water-soluble Essential Vitamins</b>			
<b>Vitamin</b>	<b>Function</b>	<b>Deficiencies Can Lead To</b>	<b>Sources</b>
Vitamin B <sub>7</sub> (Biotin)	Used in energy and amino acid metabolism, fat synthesis, and fat breakdown; helps the body use blood sugar	Hair loss, dermatitis, depression, numbness and tingling in the extremities; neuromuscular disorders	Meat, eggs, legumes and other vegetables
Vitamin B <sub>9</sub> (Folic acid)	Assists the normal development of cells, especially during fetal development; helps metabolize nucleic and amino acids	Deficiency during pregnancy is associated with birth defects, such as neural tube defects and anemia	Leafy green vegetables, whole wheat, fruits, nuts, legumes
Vitamin B <sub>12</sub> (Cobalamin)	Maintains healthy nervous system and assists with blood cell formation; coenzyme in nucleic acid metabolism	Anemia, neurological disorders, numbness, loss of balance	Meat, eggs, animal products

Water-soluble Essential Vitamins			
Vitamin	Function	Deficiencies Can Lead To	Sources
Vitamin C (Ascorbic acid)	Helps maintain connective tissue: bone, cartilage, and dentin; boosts the immune system	Scurvy, which results in bleeding, hair and tooth loss; joint pain and swelling; delayed wound healing	Citrus fruits, broccoli, tomatoes, red sweet bell peppers

Fat-soluble Essential Vitamins			
Vitamin	Function	Deficiencies Can Lead To	Sources

Fat-soluble Essential Vitamins			
Vitamin	Function	Deficiencies Can Lead To	Sources
Vitamin A (Retinol)	Critical to the development of bones, teeth, and skin; helps maintain eyesight, enhances the immune system, fetal development, gene expression	Night-blindness, skin disorders, impaired immunity	Dark green leafy vegetables, yellow-orange vegetables, fruits, milk, butter
Vitamin D	Critical for calcium absorption for bone development and strength; maintains a stable nervous system; maintains a normal and strong heartbeat; helps in blood clotting	Rickets, osteomalacia, immunity	Cod liver oil, milk, egg yolk



Fat-soluble Essential Vitamins			
Vitamin	Function	Deficiencies Can Lead To	Sources
Vitamin E (Tocopherol)	Lessens oxidative damage of cells, and prevents lung damage from pollutants; vital to the immune system	Deficiency is rare; anemia, nervous system degeneration	Wheat germ oil, unrefined vegetable oils, nuts, seeds, grains
Vitamin K (Phylloquinone)	Essential to blood clotting	Bleeding and easy bruising	Leafy green vegetables, tea



A healthy diet should include a variety of

foods to ensure that needs for essential nutrients are met. (credit: Keith Weller, USDA ARS)

Minerals and Their Function in the Human Body			
Mineral	Function	Deficiencies Can Lead To	Sources
*Calcium	Needed for muscle and neuron function; heart health; builds bone and supports synthesis and function of blood cells; nerve function	Osteoporosis, rickets, muscle spasms, impaired growth	Milk, yogurt, fish, green leafy vegetables, legumes
*Chlorine	Needed for production of hydrochloric acid (HCl) in the stomach and nerve function; osmotic balance	Muscle cramps, mood disturbances, reduced appetite	Table salt

Minerals and Their Function in the Human Body			
Mineral	Function	Deficiencies Can Lead To	Sources
Copper (trace amounts)	Required component of many redox enzymes, including cytochrome c oxidase; cofactor for hemoglobin synthesis	Copper deficiency is rare	Liver, oysters, cocoa, chocolate, sesame, nuts
Iodine	Required for the synthesis of thyroid hormones	Goiter	Seafood, iodized salt, dairy products
Iron	Required for many proteins and enzymes, notably hemoglobin, to prevent anemia	Anemia, which causes poor concentration, fatigue, and poor immune function	Red meat, leafy green vegetables, fish (tuna, salmon), eggs, dried fruits, beans, whole grains

<b>Minerals and Their Function in the Human Body</b>			
<b>Mineral</b>	<b>Function</b>	<b>Deficiencies Can Lead To</b>	<b>Sources</b>
*Magnesium	Required co-factor for ATP formation; bone formation; normal membrane functions; muscle function	Mood disturbances, muscle spasms	Whole grains, leafy green vegetables
Manganese (trace amounts)	A cofactor in enzyme functions; trace amounts are required	Manganese deficiency is rare	Common in most foods
Molybdenum (trace amounts)	Acts as a cofactor for three essential enzymes in humans: sulfite oxidase, xanthine oxidase, and aldehyde oxidase	Molybdenum deficiency is rare	
*Phosphorus	A component of bones and teeth; helps regulate acid-base balance; nucleotide synthesis	Weakness, bone abnormalities, calcium loss	Milk, hard cheese, whole grains, meats

<b>Minerals and Their Function in the Human Body</b>			
<b>Mineral</b>	<b>Function</b>	<b>Deficiencies Can Lead To</b>	<b>Sources</b>
*Potassium	Vital for muscles, heart, and nerve function	Cardiac rhythm disturbance, muscle weakness	Legumes, potato skin, tomatoes, bananas
Selenium (trace amounts)	A cofactor essential to activity of antioxidant enzymes like glutathione peroxidase; trace amounts are required	Selenium deficiency is rare	Common in most foods
*Sodium	Systemic electrolyte required for many functions; acid-base balance; water balance; nerve function	Muscle cramps, fatigue, reduced appetite	Table salt
Zinc (trace amounts)	Required for several enzymes such as carboxypeptidase, liver alcohol dehydrogenase, and carbonic anhydrase	Anemia, poor wound healing, can lead to short stature	Common in most foods

Minerals and Their Function in the Human Body			
Mineral	Function	Deficiencies Can Lead To	Sources
*Greater than 200mg/day required			

Essential Amino Acids	
Amino acids that must be consumed	Amino acids anabolized by the body
isoleucine	alanine
leucine	selenocysteine
lysine	aspartate
methionine	cysteine
phenylalanine	glutamate
tryptophan	glycine
valine	proline
histidine*	serine
threonine	tyrosine
arginine*	asparagine

<b>Essential Amino Acids</b>	
<b>Amino acids that must be consumed</b>	<b>Amino acids anabolized by the body</b>
*The human body can synthesize histidine and arginine, but not in the quantities required, especially for growing children.	

## **Food Energy and ATP**

Animals need food to obtain energy and maintain homeostasis. Homeostasis is the ability of a system to maintain a stable internal environment even in the face of external changes to the environment. For example, the normal body temperature of humans is 37°C (98.6°F). Humans maintain this temperature even when the external temperature is hot or cold. It takes energy to maintain this body temperature, and animals obtain this energy from food.

The primary source of energy for animals is carbohydrates, mainly glucose. Glucose is called the body's fuel. The digestible carbohydrates in an animal's diet are converted to glucose molecules through a series of catabolic chemical reactions.

Adenosine triphosphate, or ATP, is the primary energy currency in cells; ATP stores energy in phosphate ester bonds. ATP releases energy when the phosphodiester bonds are broken and ATP is converted to ADP and a phosphate group. ATP is produced by the oxidative reactions in the cytoplasm and mitochondrion of the cell, where carbohydrates, proteins, and fats undergo a series of metabolic reactions collectively called cellular respiration. For example, glycolysis is a series of reactions in which glucose is converted to pyruvic acid and some of its chemical potential energy is transferred to NADH and ATP.

ATP is required for all cellular functions. It is used to build the organic molecules that are required for cells and tissues; it provides energy for muscle contraction and for the transmission of electrical signals in the nervous system. When the amount of ATP is available in excess of the body's requirements, the liver uses the excess ATP and excess glucose to produce

molecules called glycogen. Glycogen is a polymeric form of glucose and is stored in the liver and skeletal muscle cells. When blood sugar drops, the liver releases glucose from stores of glycogen. Skeletal muscle converts glycogen to glucose during intense exercise. The process of converting glucose and excess ATP to glycogen and the storage of excess energy is an evolutionarily important step in helping animals deal with mobility, food shortages, and famine.

**Note:**

**Everyday Connection**

**Obesity**

Obesity is a major health concern in the United States, and there is a growing focus on reducing obesity and the diseases it may lead to, such as type-2 diabetes, cancers of the colon and breast, and cardiovascular disease. How does the food consumed contribute to obesity?

Fatty foods are calorie-dense, meaning that they have more calories per unit mass than carbohydrates or proteins. One gram of carbohydrates has four calories, one gram of protein has four calories, and one gram of fat has nine calories. Animals tend to seek lipid-rich food for their higher energy content. The signals of hunger (“time to eat”) and satiety (“time to stop eating”) are controlled in the hypothalamus region of the brain. Foods that are rich in fatty acids tend to promote satiety more than foods that are rich only in carbohydrates.

Excess carbohydrate and ATP are used by the liver to synthesize glycogen. The pyruvate produced during glycolysis is used to synthesize fatty acids. When there is more glucose in the body than required, the resulting excess pyruvate is converted into molecules that eventually result in the synthesis of fatty acids within the body. These fatty acids are stored in adipose cells—the fat cells in the mammalian body whose primary role is to store fat for later use.

It is important to note that some animals benefit from obesity. Polar bears and seals need body fat for insulation and to keep them from losing body heat during Arctic winters. When food is scarce, stored body fat provides energy for maintaining homeostasis. Fats prevent famine in mammals, allowing them to access energy when food is not available on a daily basis; fats are stored when a large kill is made or lots of food is available.



## Section Summary

Animal diet should be balanced and meet the needs of the body.

Carbohydrates, proteins, and fats are the primary components of food. Some essential nutrients are required for cellular function but cannot be produced by the animal body. These include vitamins, minerals, some fatty acids, and some amino acids. Food intake in more than necessary amounts is stored as glycogen in the liver and muscle cells, and in fat cells. Excess adipose storage can lead to obesity and serious health problems. ATP is the energy currency of the cell and is obtained from the metabolic pathways. Excess carbohydrates and energy are stored as glycogen in the body.

## Review Questions

### Exercise:

**Problem:** Which of the following statements is not true?

- a. Essential nutrients can be synthesized by the body.
- b. Vitamins are required in small quantities for bodily function.
- c. Some amino acids can be synthesized by the body, while others need to be obtained from diet.
- d. Vitamins come in two categories: fat-soluble and water-soluble.

---

### Solution:

A

### Exercise:

**Problem:** Which of the following is a water-soluble vitamin?

- a. vitamin A
  - b. vitamin E
  - c. vitamin K
  - d. vitamin C
-

**Solution:**

D

**Exercise:**

**Problem:**What is the primary fuel for the body?

- a. carbohydrates
- b. lipids
- c. protein
- d. glycogen

---

**Solution:**

A

**Exercise:**

**Problem:**Excess glucose is stored as \_\_\_\_\_.

- a. fat
- b. glucagon
- c. glycogen
- d. it is not stored in the body

---

**Solution:**

C

**Free Response**

**Exercise:**

**Problem:**What are essential nutrients?

---

**Solution:**

Essential nutrients are those nutrients that must be obtained from the diet because they cannot be produced by the body. Vitamins and minerals are examples of essential nutrients.

**Exercise:**

**Problem:**What is the role of minerals in maintaining good health?

---

**Solution:**

Minerals—such as potassium, sodium, and calcium—are required for the functioning of many cellular processes, including muscle contraction and nerve conduction. While minerals are required in trace amounts, not having minerals in the diet can be potentially harmful.

**Exercise:**

**Problem:**Discuss why obesity is a growing epidemic.

---

**Solution:**

In the United States, obesity, particularly childhood obesity, is a growing concern. Some of the contributors to this situation include sedentary lifestyles and consuming more processed foods and less fruits and vegetables. As a result, even young children who are obese can face health concerns.

**Exercise:****Problem:**

There are several nations where malnourishment is a common occurrence. What may be some of the health challenges posed by malnutrition?

---

**Solution:**

Malnutrition, often in the form of not getting enough calories or not enough of the essential nutrients, can have severe consequences. Many malnourished children have vision and dental problems, and over the years may develop many serious health problems.

## **Glossary**

essential nutrient

nutrient that cannot be synthesized by the body; it must be obtained from food

mineral

inorganic, elemental molecule that carries out important roles in the body

vitamin

organic substance necessary in small amounts to sustain life

## Digestive System Processes

By the end of this section, you will be able to:

- Describe the process of digestion
- Detail the steps involved in digestion and absorption
- Define elimination
- Explain the role of both the small and large intestines in absorption

Obtaining nutrition and energy from food is a multi-step process. For true animals, the first step is ingestion, the act of taking in food. This is followed by digestion, absorption, and elimination. In the following sections, each of these steps will be discussed in detail.

### Ingestion

The large molecules found in intact food cannot pass through the cell membranes. Food needs to be broken into smaller particles so that animals can harness the nutrients and organic molecules. The first step in this process is **ingestion**. Ingestion is the process of taking in food through the mouth. In vertebrates, the teeth, saliva, and tongue play important roles in mastication (preparing the food into bolus). While the food is being mechanically broken down, the enzymes in saliva begin to chemically process the food as well. The combined action of these processes modifies the food from large particles to a soft mass that can be swallowed and can travel the length of the esophagus.

### Digestion and Absorption

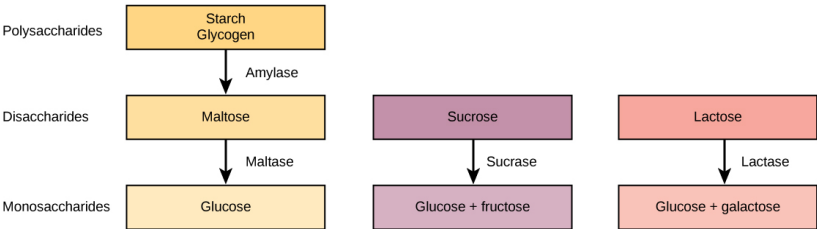
**Digestion** is the mechanical and chemical break down of food into small organic fragments. It is important to break down macromolecules into smaller fragments that are of suitable size for absorption across the digestive epithelium. Large, complex molecules of proteins, polysaccharides, and lipids must be reduced to simpler particles such as simple sugar before they can be absorbed by the digestive epithelial cells. Different organs play specific roles in the digestive process. The animal diet needs carbohydrates, protein, and fat, as well as vitamins and inorganic components for nutritional balance. How each of these components is digested is discussed in the following sections.

### Carbohydrates

The digestion of carbohydrates begins in the mouth. The salivary enzyme amylase begins the breakdown of food starches into maltose, a disaccharide. As the bolus of food travels through the esophagus to the stomach, no significant digestion of carbohydrates takes place. The esophagus produces no digestive enzymes but does produce mucous for lubrication. The acidic environment in the stomach stops the action of the amylase enzyme.

The next step of carbohydrate digestion takes place in the duodenum. Recall that the chyme from the stomach enters the duodenum and mixes with the digestive secretion from the pancreas, liver, and gallbladder. Pancreatic juices also contain amylase, which continues the breakdown of starch and glycogen into maltose, a disaccharide. The disaccharides are broken down into

monosaccharides by enzymes called **maltases**, **sucrases**, and **lactases**, which are also present in the brush border of the small intestinal wall. Maltase breaks down maltose into glucose. Other disaccharides, such as sucrose and lactose are broken down by sucrase and lactase, respectively. Sucrase breaks down sucrose (or “table sugar”) into glucose and fructose, and lactase breaks down lactose (or “milk sugar”) into glucose and galactose. The monosaccharides (glucose) thus produced are absorbed and then can be used in metabolic pathways to harness energy. The monosaccharides are transported across the intestinal epithelium into the bloodstream to be transported to the different cells in the body. The steps in carbohydrate digestion are summarized in [\[link\]](#) and [\[link\]](#).



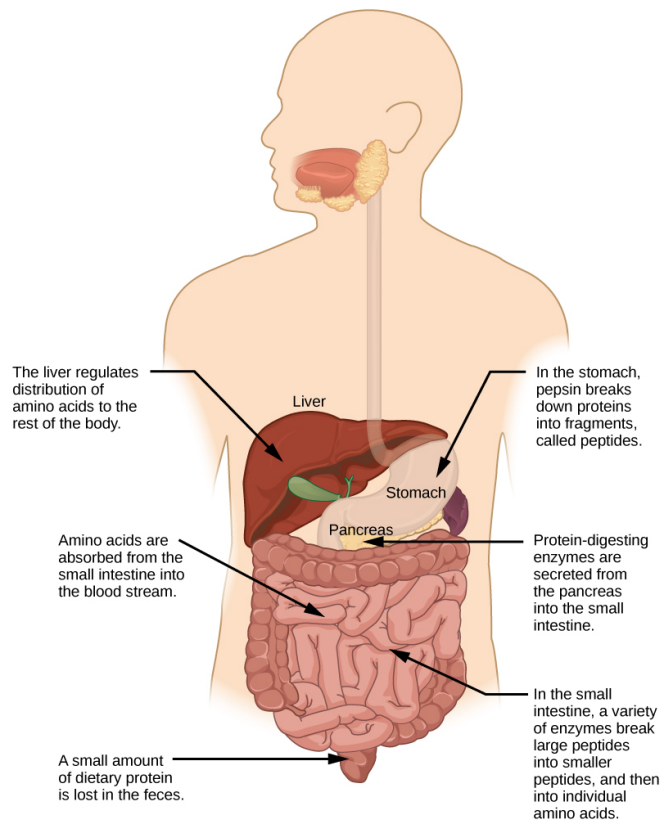
Digestion of carbohydrates is performed by several enzymes. Starch and glycogen are broken down into glucose by amylase and maltase. Sucrose (table sugar) and lactose (milk sugar) are broken down by sucrase and lactase, respectively.

Digestion of Carbohydrates				
Enzyme	Produced By	Site of Action	Substrate Acting On	End Products
Salivary amylase	Salivary glands	Mouth	Polysaccharides (Starch)	Disaccharides (maltose), oligosaccharides
Pancreatic amylase	Pancreas	Small intestine	Polysaccharides (starch)	Disaccharides (maltose), monosaccharides

Digestion of Carbohydrates				
Enzyme	Produced By	Site of Action	Substrate Acting On	End Products
Oligosaccharidases	Lining of the intestine; brush border membrane	Small intestine	Disaccharides	Monosaccharides (e.g., glucose, fructose, galactose)

## Protein

A large part of protein digestion takes place in the stomach. The enzyme pepsin plays an important role in the digestion of proteins by breaking down the intact protein to peptides, which are short chains of four to nine amino acids. In the duodenum, other enzymes—**trypsin**, **elastase**, and **chymotrypsin**—act on the peptides reducing them to smaller peptides. Trypsin elastase, carboxypeptidase, and chymotrypsin are produced by the pancreas and released into the duodenum where they act on the chyme. Further breakdown of peptides to single amino acids is aided by enzymes called peptidases (those that break down peptides). Specifically, **carboxypeptidase**, **dipeptidase**, and **aminopeptidase** play important roles in reducing the peptides to free amino acids. The amino acids are absorbed into the bloodstream through the small intestines. The steps in protein digestion are summarized in [\[link\]](#) and [\[link\]](#).



Protein digestion is a multistep process that begins in the stomach and continues through the intestines.

Digestion of Protein				
Enzyme	Produced By	Site of Action	Substrate Acting On	End Products
Pepsin	Stomach chief cells	Stomach	Proteins	Peptides
Trypsin Elastase Chymotrypsin	Pancreas	Small intestine	Proteins	Peptides



Digestion of Protein				
Enzyme	Produced By	Site of Action	Substrate Acting On	End Products
Carboxypeptidase	Pancreas	Small intestine	Peptides	Amino acids and peptides
Aminopeptidase Dipeptidase	Lining of intestine	Small intestine	Peptides	Amino acids

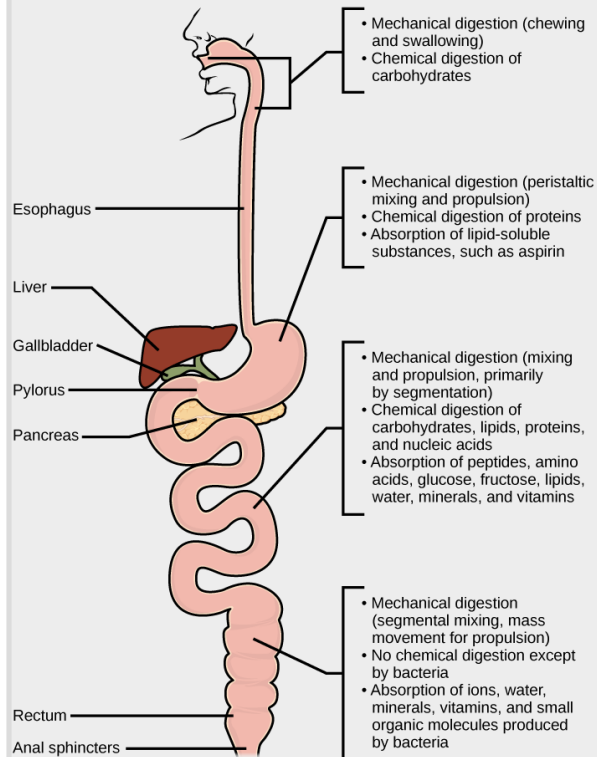
## Lipids

Lipid digestion begins in the stomach with the aid of lingual lipase and gastric lipase. However, the bulk of lipid digestion occurs in the small intestine due to pancreatic lipase. When chyme enters the duodenum, the hormonal responses trigger the release of bile, which is produced in the liver and stored in the gallbladder. Bile aids in the digestion of lipids, primarily triglycerides by emulsification. Emulsification is a process in which large lipid globules are broken down into several small lipid globules. These small globules are more widely distributed in the chyme rather than forming large aggregates. Lipids are hydrophobic substances: in the presence of water, they will aggregate to form globules to minimize exposure to water. Bile contains bile salts, which are amphipathic, meaning they contain hydrophobic and hydrophilic parts. Thus, the bile salts hydrophilic side can interface with water on one side and the hydrophobic side interfaces with lipids on the other. By doing so, bile salts emulsify large lipid globules into small lipid globules.

Why is emulsification important for digestion of lipids? Pancreatic juices contain enzymes called lipases (enzymes that break down lipids). If the lipid in the chyme aggregates into large globules, very little surface area of the lipids is available for the lipases to act on, leaving lipid digestion incomplete. By forming an emulsion, bile salts increase the available surface area of the lipids many fold. The pancreatic lipases can then act on the lipids more efficiently and digest them, as detailed in [\[link\]](#). Lipases break down the lipids into fatty acids and glycerides. These molecules can pass through the plasma membrane of the cell and enter the epithelial cells of the intestinal lining. The bile salts surround long-chain fatty acids and monoglycerides forming tiny spheres called micelles. The micelles move into the brush border of the small intestine absorptive cells where the long-chain fatty acids and monoglycerides diffuse out of the micelles into the absorptive cells leaving the micelles behind in the chyme. The long-chain fatty acids and monoglycerides recombine in the absorptive cells to form triglycerides, which aggregate into globules and become coated with proteins. These large spheres are called **chylomicrons**. Chylomicrons contain triglycerides, cholesterol, and other lipids and have proteins on their surface. The surface is also composed of the hydrophilic phosphate "heads" of phospholipids. Together, they enable the chylomicron to move in an aqueous environment without exposing the lipids to water. Chylomicrons leave the absorptive cells via exocytosis. Chylomicrons enter the lymphatic vessels, and then enter the blood in the subclavian vein.



**Note:**  
**Art Connection**



Mechanical and chemical digestion of food takes place in many steps, beginning in the mouth and ending in the rectum.

Which of the following statements about digestive processes is true?

- a. Amylase, maltase, and lactase in the mouth digest carbohydrates.
- b. Trypsin and lipase in the stomach digest protein.
- c. Bile emulsifies lipids in the small intestine.
- d. No food is absorbed until the small intestine.

## Elimination

The final step in digestion is the elimination of undigested food content and waste products. The undigested food material enters the colon, where most of the water is reabsorbed. Recall that the colon is also home to the microflora called “intestinal flora” that aid in the digestion process. The semi-solid waste is moved through the colon by peristaltic movements of the muscle and is stored in the rectum. As the rectum expands in response to storage of fecal matter, it triggers the neural signals required to set up the urge to eliminate. The solid waste is eliminated through the anus using peristaltic movements of the rectum.

## Common Problems with Elimination

Diarrhea and constipation are some of the most common health concerns that affect digestion. Constipation is a condition where the feces are hardened because of excess water removal in the colon. In contrast, if enough water is not removed from the feces, it results in diarrhea. Many bacteria, including the ones that cause cholera, affect the proteins involved in water reabsorption in the colon and result in excessive diarrhea.

## Emesis

Emesis, or vomiting, is elimination of food by forceful expulsion through the mouth. It is often in response to an irritant that affects the digestive tract, including but not limited to viruses, bacteria, emotions, sights, and food poisoning. This forceful expulsion of the food is due to the strong contractions produced by the stomach muscles. The process of emesis is regulated by the medulla.

## Section Summary

Digestion begins with ingestion, where the food is taken in the mouth. Digestion and absorption take place in a series of steps with special enzymes playing important roles in digesting carbohydrates, proteins, and lipids. Elimination describes removal of undigested food contents and waste products from the body. While most absorption occurs in the small intestines, the large intestine is responsible for the final removal of water that remains after the absorptive process of the small intestines. The cells that line the large intestine absorb some vitamins as well as any leftover salts and water. The large intestine (colon) is also where feces is formed.

## Art Connections

### Exercise:

**Problem:** [\[link\]](#) Which of the following statements about digestive processes is true?

- a. Amylase, maltase and lactase in the mouth digest carbohydrates.
- b. Trypsin and lipase in the stomach digest protein.
- c. Bile emulsifies lipids in the small intestine.
- d. No food is absorbed until the small intestine.

---

### Solution:

[\[link\]](#) C

## Review Questions

### Exercise:

**Problem:** Where does the majority of protein digestion take place?

- a. stomach
- b. duodenum
- c. mouth
- d. jejunum

---

**Solution:**

A

**Exercise:**

**Problem:** Lipases are enzymes that break down \_\_\_\_\_.

- a. disaccharides
- b. lipids
- c. proteins
- d. cellulose

---

**Solution:**

B

## Free Response

**Exercise:**

**Problem:** Explain why some dietary lipid is a necessary part of a balanced diet.

---

**Solution:**

Lipids add flavor to food and promote a sense of satiety or fullness. Fatty foods are sources of high energy; one gram of lipid contains nine calories. Lipids are also required in the diet to aid the absorption of lipid-soluble vitamins and for the production of lipid-soluble hormones.

## Glossary

aminopeptidase

protease that breaks down peptides to single amino acids; secreted by the brush border of small intestine

carboxypeptidase

protease that breaks down peptides to single amino acids; secreted by the brush border of the small intestine

chylomicron  
small lipid globule

chymotrypsin  
pancreatic protease

digestion  
mechanical and chemical break down of food into small organic fragments

dipeptidase  
protease that breaks down peptides to single amino acids; secreted by the brush border of small intestine

elastase  
pancreatic protease

ingestion  
act of taking in food

lactase  
enzyme that breaks down lactose into glucose and galactose

maltase  
enzyme that breaks down maltose into glucose

sucrase  
enzyme that breaks down sucrose into glucose and fructose

trypsin  
pancreatic protease that breaks down protein

## Digestive System Regulation

By the end of this section, you will be able to:

- Discuss the role of neural regulation in digestive processes
- Explain how hormones regulate digestion

The brain is the control center for the sensation of hunger and satiety. The functions of the digestive system are regulated through neural and hormonal responses.

## Neural Responses to Food

In reaction to the smell, sight, or thought of food, like that shown in [\[link\]](#), the first response is that of salivation. The salivary glands secrete more saliva in response to stimulation by the autonomic nervous system triggered by food in preparation for digestion. Simultaneously, the stomach begins to produce hydrochloric acid to digest the food. Recall that the peristaltic movements of the esophagus and other organs of the digestive tract are under the control of the brain. The brain prepares these muscles for movement as well. When the stomach is full, the part of the brain that detects satiety signals fullness. There are three overlapping phases of gastric control—the cephalic phase, the gastric phase, and the intestinal phase—each requires many enzymes and is under neural control as well.



Seeing a plate of food triggers the secretion of saliva in the mouth and the production of HCL in the stomach.  
(credit: Kelly Bailey)

## Digestive Phases

The response to food begins even before food enters the mouth. The first phase of ingestion, called the **cephalic phase**, is controlled by the neural response to the stimulus provided by food. All aspects—such as sight, sense, and smell—trigger the neural responses resulting in salivation and secretion of gastric juices. The gastric and salivary secretion in the cephalic phase can also take place due to the thought of food. Right now, if you think about a piece of chocolate or a crispy potato chip, the increase in salivation is a cephalic phase response to the thought. The central nervous system prepares the stomach to receive food.

The **gastric phase** begins once the food arrives in the stomach. It builds on the stimulation provided during the cephalic phase. Gastric acids and enzymes process the ingested materials. The gastric phase is stimulated by



(1) distension of the stomach, (2) a decrease in the pH of the gastric contents, and (3) the presence of undigested material. This phase consists of local, hormonal, and neural responses. These responses stimulate secretions and powerful contractions.

The **intestinal phase** begins when chyme enters the small intestine triggering digestive secretions. This phase controls the rate of gastric emptying. In addition to gastrin emptying, when chyme enters the small intestine, it triggers other hormonal and neural events that coordinate the activities of the intestinal tract, pancreas, liver, and gallbladder.

## Hormonal Responses to Food

The **endocrine system** controls the response of the various glands in the body and the release of hormones at the appropriate times.

One of the important factors under hormonal control is the stomach acid environment. During the gastric phase, the hormone **gastrin** is secreted by G cells in the stomach in response to the presence of proteins. Gastrin stimulates the release of stomach acid, or hydrochloric acid (HCl) which aids in the digestion of the proteins. However, when the stomach is emptied, the acidic environment need not be maintained and a hormone called **somatostatin** stops the release of hydrochloric acid. This is controlled by a negative feedback mechanism.

In the duodenum, digestive secretions from the liver, pancreas, and gallbladder play an important role in digesting chyme during the intestinal phase. In order to neutralize the acidic chyme, a hormone called **secretin** stimulates the pancreas to produce alkaline bicarbonate solution and deliver it to the duodenum. Secretin acts in tandem with another hormone called **cholecystokinin** (CCK). Not only does CCK stimulate the pancreas to produce the requisite pancreatic juices, it also stimulates the gallbladder to release bile into the duodenum.

**Note:**

## Link to Learning



Visit [this website](#) to learn more about the endocrine system. Review the text and watch the animation of how control is implemented in the endocrine system.

Another level of hormonal control occurs in response to the composition of food. Foods high in lipids take a long time to digest. A hormone called **gastric inhibitory peptide** is secreted by the small intestine to slow down the peristaltic movements of the intestine to allow fatty foods more time to be digested and absorbed.

Understanding the hormonal control of the digestive system is an important area of ongoing research. Scientists are exploring the role of each hormone in the digestive process and developing ways to target these hormones. Advances could lead to knowledge that may help to battle the obesity epidemic.

## Section Summary

The brain and the endocrine system control digestive processes. The brain controls the responses of hunger and satiety. The endocrine system controls the release of hormones and enzymes required for digestion of food in the digestive tract.

## Review Questions

### Exercise:

**Problem:**

Which hormone controls the release of bile from the gallbladder

- a. pepsin
- b. amylase
- c. CCK
- d. gastrin

---

**Solution:**

C

**Exercise:**

**Problem:** Which hormone stops acid secretion in the stomach?

- a. gastrin
- b. somatostatin
- c. gastric inhibitory peptide
- d. CCK

---

**Solution:**

B

**Free Response****Exercise:**

**Problem:** Describe how hormones regulate digestion.

---

**Solution:**

Hormones control the different digestive enzymes that are secreted in the stomach and the intestine during the process of digestion and absorption. For example, the hormone gastrin stimulates stomach acid secretion in response to food intake. The hormone somatostatin stops the release of stomach acid.

**Exercise:**

**Problem:**

Describe one or more scenarios where loss of hormonal regulation of digestion can lead to diseases.

---

**Solution:**

There are many cases where loss of hormonal regulation can lead to illnesses. For example, the bilirubin produced by the breakdown of red blood cells is converted to bile by the liver. When there is malfunction of this process, there is excess bilirubin in the blood and bile levels are low. As a result, the body struggles with dealing with fatty food. This is why a patient suffering from jaundice is asked to eat a diet with almost zero fat.

## **Glossary**

cephalic phase

first phase of digestion, controlled by the neural response to the stimulus provided by food

cholecystokinin

hormone that stimulates the contraction of the gallbladder to release bile

endocrine system

system that controls the response of the various glands in the body and the release of hormones at the appropriate times

gastric inhibitory peptide

hormone secreted by the small intestine in the presence of fatty acids and sugars; it also inhibits acid production and peristalsis in order to slow down the rate at which food enters the small intestine

gastric phase

digestive phase beginning once food enters the stomach; gastric acids and enzymes process the ingested materials

gastrin

hormone which stimulates hydrochloric acid secretion in the stomach

intestinal phase

third digestive phase; begins when chyme enters the small intestine triggering digestive secretions and controlling the rate of gastric emptying

secretin

hormone which stimulates sodium bicarbonate secretion in the small intestine

somatostatin

hormone released to stop acid secretion when the stomach is empty

## Introduction

class="introduction"

Improvements in the design of prostheses have allowed for a wider range of activities in recipients.

(credit: modification of work by Stuart Grout)



The muscular and skeletal systems provide support to the body and allow for a wide range of movement. The bones of the skeletal system protect the

body's internal organs and support the weight of the body. The muscles of the muscular system contract and pull on the bones, allowing for movements as diverse as standing, walking, running, and grasping items.

Injury or disease affecting the musculoskeletal system can be very debilitating. In humans, the most common musculoskeletal diseases worldwide are caused by malnutrition. Ailments that affect the joints are also widespread, such as arthritis, which can make movement difficult and—in advanced cases—completely impair mobility. In severe cases in which the joint has suffered extensive damage, joint replacement surgery may be needed.

Progress in the science of prosthesis design has resulted in the development of artificial joints, with joint replacement surgery in the hips and knees being the most common. Replacement joints for shoulders, elbows, and fingers are also available. Even with this progress, there is still room for improvement in the design of prostheses. The state-of-the-art prostheses have limited durability and therefore wear out quickly, particularly in young or active individuals. Current research is focused on the use of new materials, such as carbon fiber, that may make prostheses more durable.

## Types of Skeletal Systems

By the end of this section, you will be able to:

- Discuss the different types of skeletal systems
- Explain the role of the human skeletal system
- Compare and contrast different skeletal systems

A skeletal system is necessary to support the body, protect internal organs, and allow for the movement of an organism. There are three different skeleton designs that fulfill these functions: hydrostatic skeleton, exoskeleton, and endoskeleton.

### Hydrostatic Skeleton

A **hydrostatic skeleton** is a skeleton formed by a fluid-filled compartment within the body, called the coelom. The organs of the coelom are supported by the aqueous fluid, which also resists external compression. This compartment is under hydrostatic pressure because of the fluid and supports the other organs of the organism. This type of skeletal system is found in soft-bodied animals such as sea anemones, earthworms, Cnidaria, and other invertebrates ([\[link\]](#)).



The skeleton of the red-knobbed sea star (*Protoreaster linckii*) is an example of a hydrostatic skeleton.



(credit: “Amada44”/Wikimedia Commons)

Movement in a hydrostatic skeleton is provided by muscles that surround the coelom. The muscles in a hydrostatic skeleton contract to change the shape of the coelom; the pressure of the fluid in the coelom produces movement. For example, earthworms move by waves of muscular contractions of the skeletal muscle of the body wall hydrostatic skeleton, called peristalsis, which alternately shorten and lengthen the body. Lengthening the body extends the anterior end of the organism. Most organisms have a mechanism to fix themselves in the substrate. Shortening the muscles then draws the posterior portion of the body forward. Although a hydrostatic skeleton is well-suited to invertebrate organisms such as earthworms and some aquatic organisms, it is not an efficient skeleton for terrestrial animals.

## Exoskeleton

An **exoskeleton** is an external skeleton that consists of a hard encasement on the surface of an organism. For example, the shells of crabs and insects are exoskeletons ([\[link\]](#)). This skeleton type provides defence against predators, supports the body, and allows for movement through the contraction of attached muscles. As with vertebrates, muscles must cross a joint inside the exoskeleton. Shortening of the muscle changes the relationship of the two segments of the exoskeleton. Arthropods such as crabs and lobsters have exoskeletons that consist of 30–50 percent chitin, a polysaccharide derivative of glucose that is a strong but flexible material. Chitin is secreted by the epidermal cells. The exoskeleton is further strengthened by the addition of calcium carbonate in organisms such as the lobster. Because the exoskeleton is acellular, arthropods must periodically shed their exoskeletons because the exoskeleton does not grow as the organism grows.



Muscles attached to the exoskeleton of the Halloween crab (*Gecarcinus quadratus*) allow it to move.

## Endoskeleton

An **endoskeleton** is a skeleton that consists of hard, mineralized structures located within the soft tissue of organisms. An example of a primitive endoskeletal structure is the spicules of sponges. The bones of vertebrates are composed of tissues, whereas sponges have no true tissues ([\[link\]](#)). Endoskeletons provide support for the body, protect internal organs, and allow for movement through contraction of muscles attached to the skeleton.



The skeletons of humans and horses are examples of endoskeletons. (credit: Ross Murphy)

The human skeleton is an endoskeleton that consists of 206 bones in the adult. It has five main functions: providing support to the body, storing minerals and lipids, producing blood cells, protecting internal organs, and allowing for movement. The skeletal system in vertebrates is divided into the axial skeleton (which consists of the skull, vertebral column, and rib cage), and the appendicular skeleton (which consists of the shoulders, limb bones, the pectoral girdle, and the pelvic girdle).

**Note:**

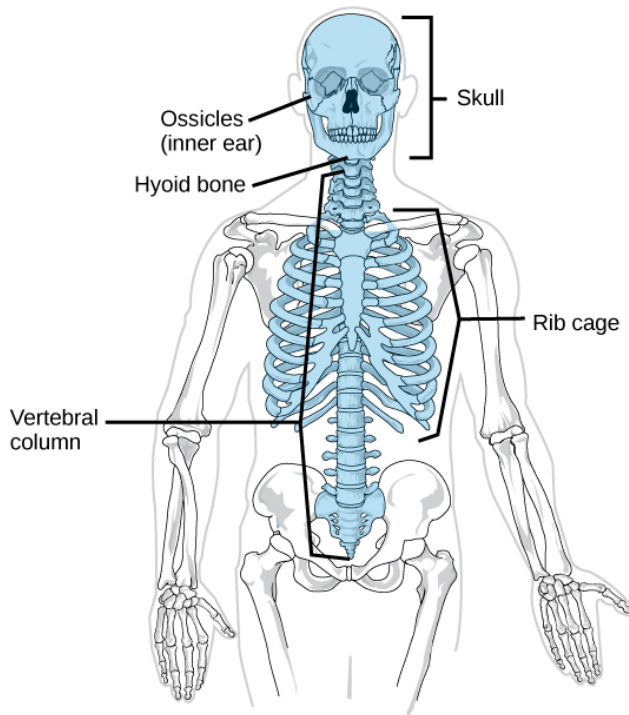
[Link to Learning](#)



Visit the [interactive body](#) site to build a virtual skeleton: select "skeleton" and click through the activity to place each bone.

## Human Axial Skeleton

The **axial skeleton** forms the central axis of the body and includes the bones of the skull, ossicles of the middle ear, hyoid bone of the throat, vertebral column, and the thoracic cage (ribcage) ([link](#)). The function of the axial skeleton is to provide support and protection for the brain, the spinal cord, and the organs in the ventral body cavity. It provides a surface for the attachment of muscles that move the head, neck, and trunk, performs respiratory movements, and stabilizes parts of the appendicular skeleton.

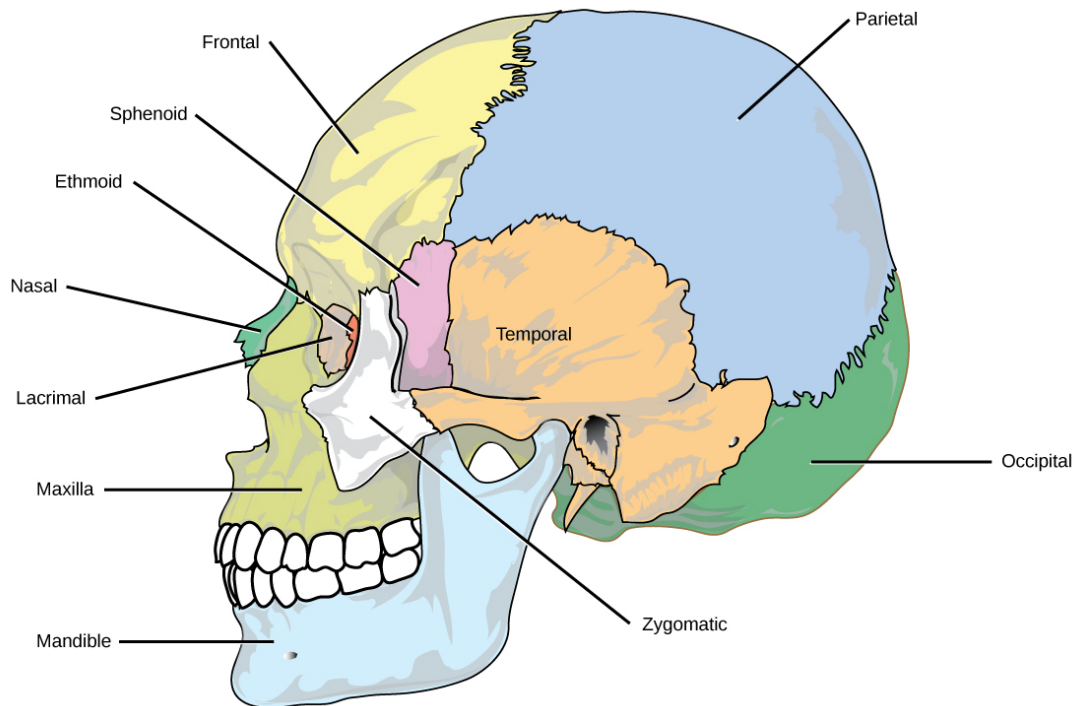


The axial skeleton consists of the bones of the skull, ossicles of the middle ear, hyoid bone, vertebral column, and rib cage. (credit: modification of work by Mariana Ruiz Villareal)

## The Skull

The bones of the **skull** support the structures of the face and protect the brain. The skull consists of 22 bones, which are divided into two categories: cranial bones and facial bones. The **cranial bones** are eight bones that form the cranial cavity, which encloses the brain and serves as an attachment site for the muscles of the head and neck. The eight cranial bones are the frontal bone, two parietal bones, two temporal bones, occipital bone, sphenoid bone, and the ethmoid bone. Although the bones developed separately in

the embryo and fetus, in the adult, they are tightly fused with connective tissue and adjoining bones do not move ([link](#)).

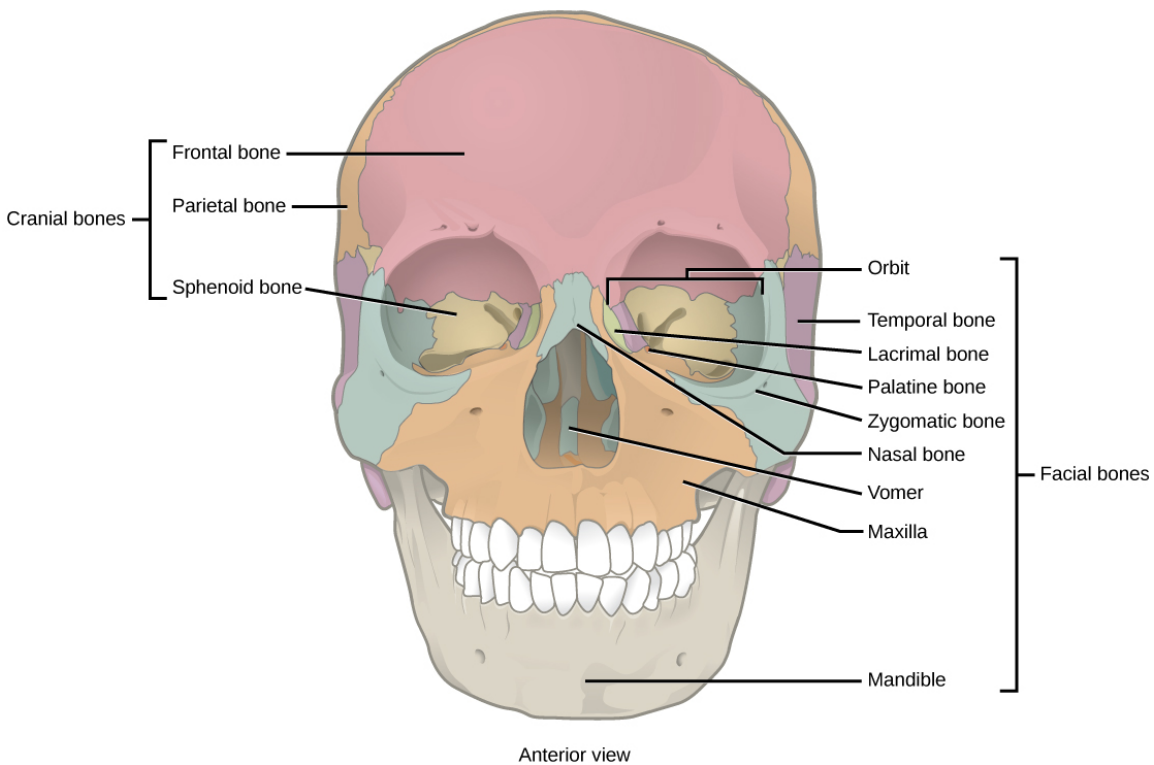


The bones of the skull support the structures of the face and protect the brain. (credit: modification of work by Mariana Ruiz Villareal)

The **auditory ossicles** of the middle ear transmit sounds from the air as vibrations to the fluid-filled cochlea. The auditory ossicles consist of three bones each: the malleus, incus, and stapes. These are the smallest bones in the body and are unique to mammals.

Fourteen **facial bones** form the face, provide cavities for the sense organs (eyes, mouth, and nose), protect the entrances to the digestive and respiratory tracts, and serve as attachment points for facial muscles. The 14 facial bones are the nasal bones, the maxillary bones, zygomatic bones, palatine, vomer, lacrimal bones, the inferior nasal conchae, and the

mandible. All of these bones occur in pairs except for the mandible and the vomer ([\[link\]](#)).



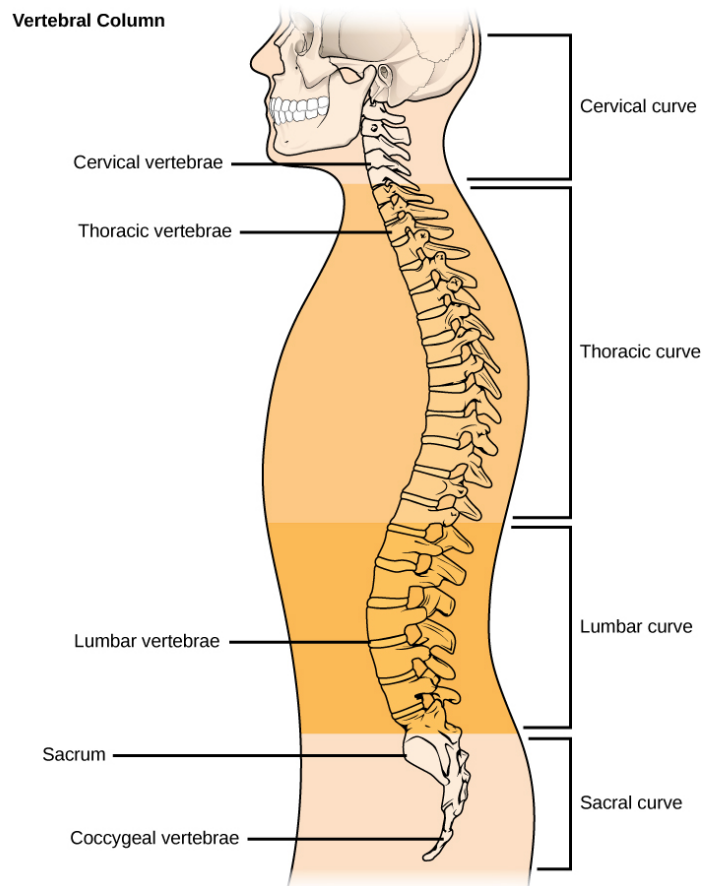
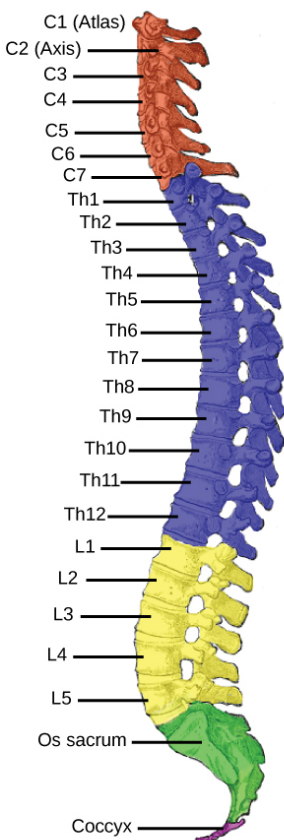
The cranial bones, including the frontal, parietal, and sphenoid bones, cover the top of the head. The facial bones of the skull form the face and provide cavities for the eyes, nose, and mouth.

Although it is not found in the skull, the hyoid bone is considered a component of the axial skeleton. The **hyoid bone** lies below the mandible in the front of the neck. It acts as a movable base for the tongue and is connected to muscles of the jaw, larynx, and tongue. The mandible articulates with the base of the skull. The mandible controls the opening to the airway and gut. In animals with teeth, the mandible brings the surfaces of the teeth in contact with the maxillary teeth.



## The Vertebral Column

The **vertebral column**, or spinal column, surrounds and protects the spinal cord, supports the head, and acts as an attachment point for the ribs and muscles of the back and neck. The adult vertebral column comprises 26 bones: the 24 vertebrae, the sacrum, and the coccyx bones. In the adult, the sacrum is typically composed of five vertebrae that fuse into one. The coccyx is typically 3–4 vertebrae that fuse into one. Around the age of 70, the sacrum and the coccyx may fuse together. We begin life with approximately 33 vertebrae, but as we grow, several vertebrae fuse together. The adult vertebrae are further divided into the 7 cervical vertebrae, 12 thoracic vertebrae, and 5 lumbar vertebrae ([\[link\]](#)).



(a) The vertebral column consists of seven cervical vertebrae (C1–7) twelve thoracic vertebrae (Th1–12), five lumbar



vertebrae (L1–5), the os sacrum, and the coccyx. (b) Spinal curves increase the strength and flexibility of the spine. (credit a: modification of work by Uwe Gille based on original work by Gray's Anatomy; credit b: modification of work by NCI, NIH)

Each vertebral body has a large hole in the center through which the nerves of the spinal cord pass. There is also a notch on each side through which the spinal nerves, which serve the body at that level, can exit from the spinal cord. The vertebral column is approximately 71 cm (28 inches) in adult male humans and is curved, which can be seen from a side view. The names of the spinal curves correspond to the region of the spine in which they occur. The thoracic and sacral curves are concave (curve inwards relative to the front of the body) and the cervical and lumbar curves are convex (curve outwards relative to the front of the body). The arched curvature of the vertebral column increases its strength and flexibility, allowing it to absorb shocks like a spring ([link](#)).

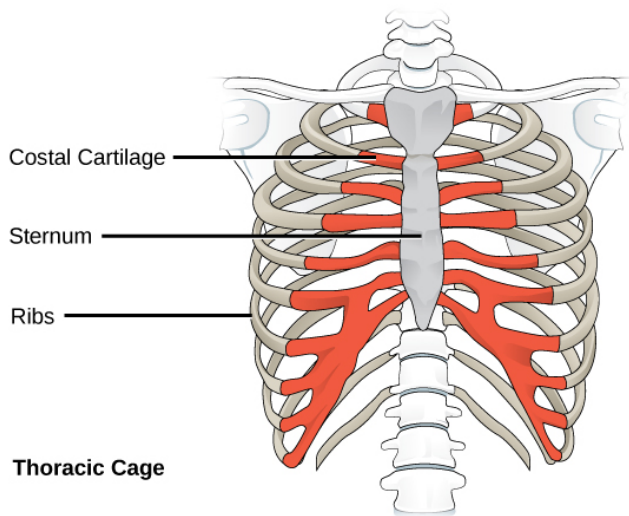
**Intervertebral discs** composed of fibrous cartilage lie between adjacent vertebral bodies from the second cervical vertebra to the sacrum. Each disc is part of a joint that allows for some movement of the spine and acts as a cushion to absorb shocks from movements such as walking and running. Intervertebral discs also act as ligaments to bind vertebrae together. The inner part of discs, the nucleus pulposus, hardens as people age and becomes less elastic. This loss of elasticity diminishes its ability to absorb shocks.

## The Thoracic Cage

The **thoracic cage**, also known as the ribcage, is the skeleton of the chest, and consists of the ribs, sternum, thoracic vertebrae, and costal cartilages ([link](#)). The thoracic cage encloses and protects the organs of the thoracic cavity, including the heart and lungs. It also provides support for the shoulder girdles and upper limbs, and serves as the attachment point for the

diaphragm, muscles of the back, chest, neck, and shoulders. Changes in the volume of the thorax enable breathing.

The **sternum**, or breastbone, is a long, flat bone located at the anterior of the chest. It is formed from three bones that fuse in the adult. The **ribs** are 12 pairs of long, curved bones that attach to the thoracic vertebrae and curve toward the front of the body, forming the ribcage. Costal cartilages connect the anterior ends of the ribs to the sternum, with the exception of rib pairs 11 and 12, which are free-floating ribs.

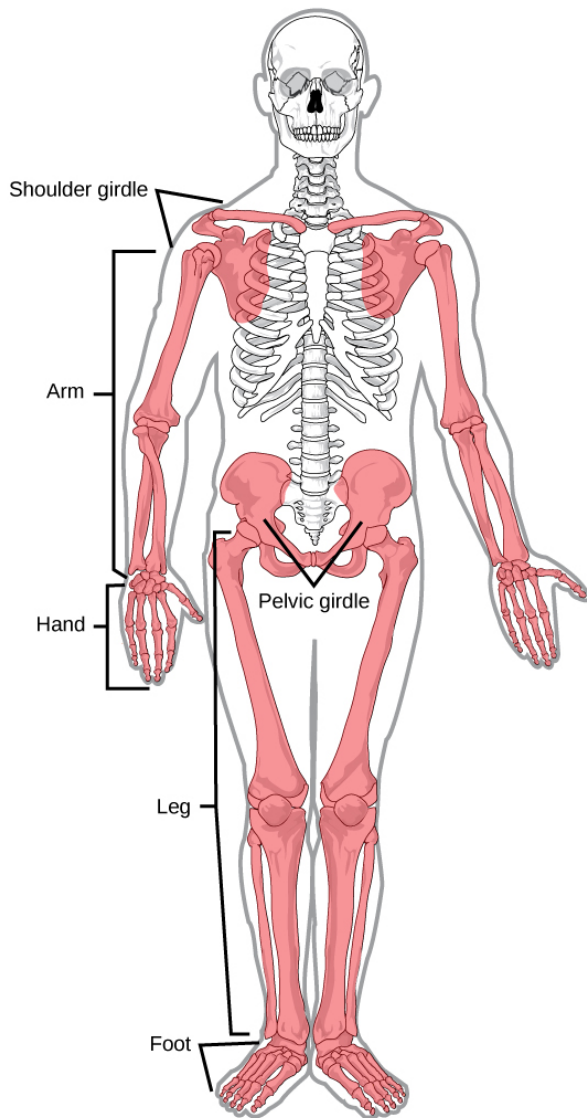


The thoracic cage, or rib cage, protects the heart and the lungs.  
(credit: modification of work by NCI, NIH)

## Human Appendicular Skeleton

The **appendicular skeleton** is composed of the bones of the upper limbs (which function to grasp and manipulate objects) and the lower limbs (which permit locomotion). It also includes the pectoral girdle, or shoulder

girdle, that attaches the upper limbs to the body, and the pelvic girdle that attaches the lower limbs to the body ([\[link\]](#)).

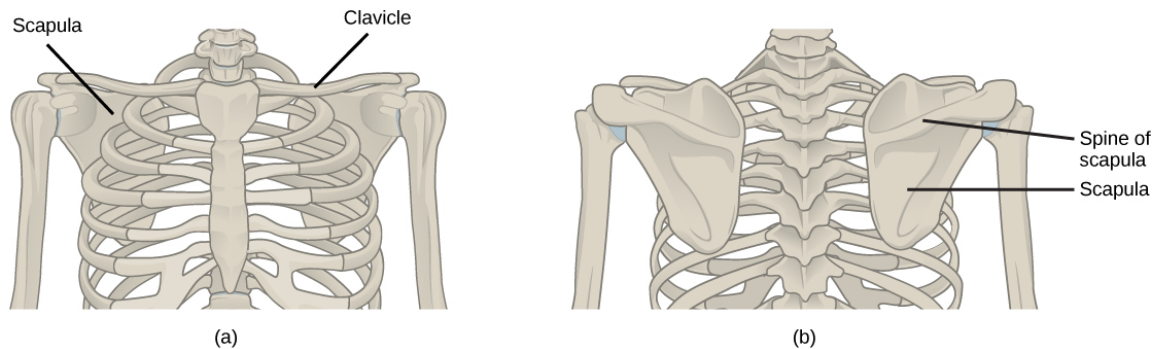


The appendicular skeleton is composed of the bones of the pectoral limbs (arm, forearm, hand), the pelvic limbs (thigh, leg, foot), the pectoral girdle, and the pelvic girdle. (credit:

modification of work by  
Mariana Ruiz Villareal)

## The Pectoral Girdle

The **pectoral girdle** bones provide the points of attachment of the upper limbs to the axial skeleton. The human pectoral girdle consists of the clavicle (or collarbone) in the anterior, and the scapula (or shoulder blades) in the posterior ([link](#)).



(a) The pectoral girdle in primates consists of the clavicles and scapulae. (b) The posterior view reveals the spine of the scapula to which muscle attaches.

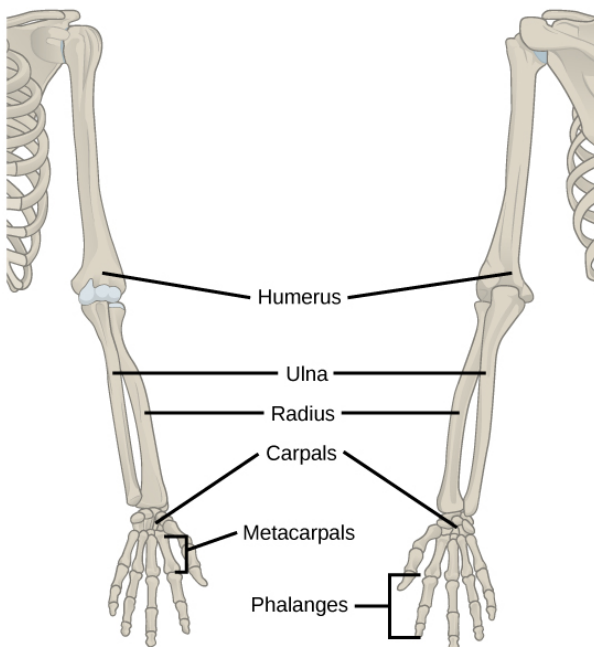
The **clavicles** are S-shaped bones that position the arms on the body. The clavicles lie horizontally across the front of the thorax (chest) just above the first rib. These bones are fairly fragile and are susceptible to fractures. For example, a fall with the arms outstretched causes the force to be transmitted to the clavicles, which can break if the force is excessive. The clavicle articulates with the sternum and the scapula.

The **scapulae** are flat, triangular bones that are located at the back of the pectoral girdle. They support the muscles crossing the shoulder joint. A

ridge, called the spine, runs across the back of the scapula and can easily be felt through the skin ([\[link\]](#)). The spine of the scapula is a good example of a bony protrusion that facilitates a broad area of attachment for muscles to bone.

## The Upper Limb

The upper limb contains 30 bones in three regions: the arm (shoulder to elbow), the forearm (ulna and radius), and the wrist and hand ([\[link\]](#)).



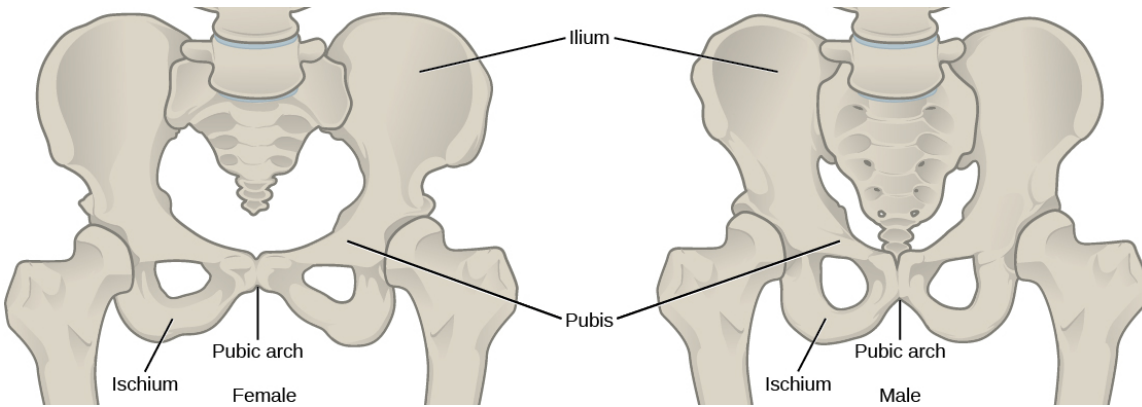
The upper limb consists of the humerus of the upper arm, the radius and ulna of the forearm, eight bones of the carpus, five bones of the metacarpus, and 14 bones of the phalanges.

An **articulation** is any place at which two bones are joined. The **humerus** is the largest and longest bone of the upper limb and the only bone of the arm. It articulates with the scapula at the shoulder and with the forearm at the elbow. The **forearm** extends from the elbow to the wrist and consists of two bones: the ulna and the radius. The **radius** is located along the lateral (thumb) side of the forearm and articulates with the humerus at the elbow. The **ulna** is located on the medial aspect (pinky-finger side) of the forearm. It is longer than the radius. The ulna articulates with the humerus at the elbow. The radius and ulna also articulate with the carpal bones and with each other, which in vertebrates enables a variable degree of rotation of the carpus with respect to the long axis of the limb. The hand includes the eight bones of the **carpus** (wrist), the five bones of the **metacarpus** (palm), and the 14 bones of the **phalanges** (digits). Each digit consists of three phalanges, except for the thumb, when present, which has only two.

## The Pelvic Girdle

The **pelvic girdle** attaches to the lower limbs of the axial skeleton. Because it is responsible for bearing the weight of the body and for locomotion, the pelvic girdle is securely attached to the axial skeleton by strong ligaments. It also has deep sockets with robust ligaments to securely attach the femur to the body. The pelvic girdle is further strengthened by two large hip bones. In adults, the hip bones, or **coxal bones**, are formed by the fusion of three pairs of bones: the ilium, ischium, and pubis. The pelvis joins together in the anterior of the body at a joint called the pubic symphysis and with the bones of the sacrum at the posterior of the body.

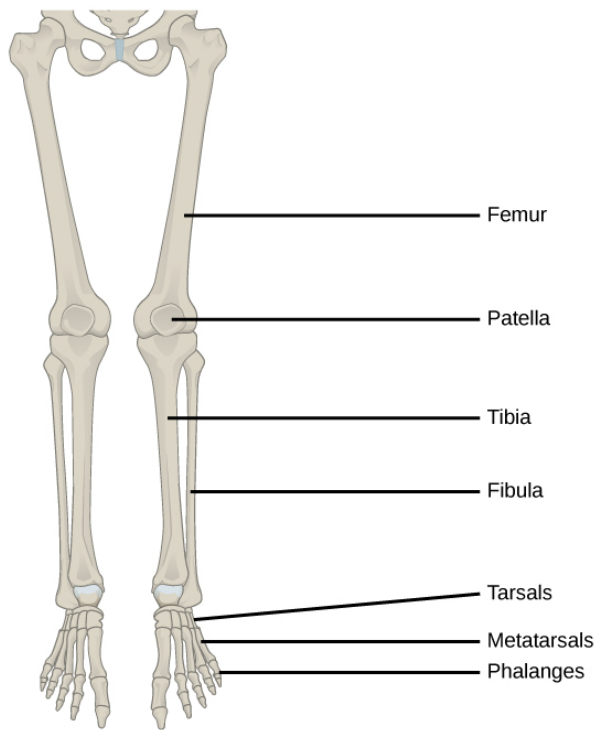
The female pelvis is slightly different from the male pelvis. Over generations of evolution, females with a wider pubic angle and larger diameter pelvic canal reproduced more successfully. Therefore, their offspring also had pelvic anatomy that enabled successful childbirth ([link](#)).



To adapt to reproductive fitness, the (a) female pelvis is lighter, wider, shallower, and has a broader angle between the pubic bones than (b) the male pelvis.

## The Lower Limb

The **lower limb** consists of the thigh, the leg, and the foot. The bones of the lower limb are the femur (thigh bone), patella (kneecap), tibia and fibula (bones of the leg), tarsals (bones of the ankle), and metatarsals and phalanges (bones of the foot) ([link](#)). The bones of the lower limbs are thicker and stronger than the bones of the upper limbs because of the need to support the entire weight of the body and the resulting forces from locomotion. In addition to evolutionary fitness, the bones of an individual will respond to forces exerted upon them.



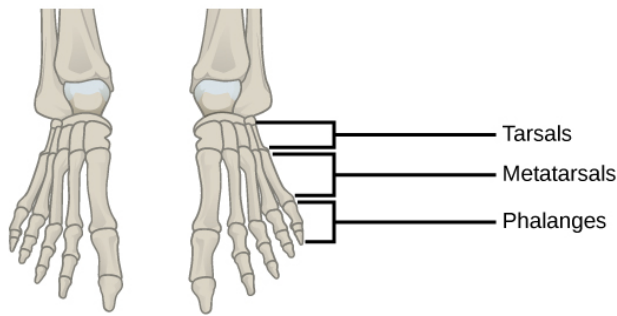
The lower limb consists of the thigh (femur), kneecap (patella), leg (tibia and fibula), ankle (tarsals), and foot (metatarsals and phalanges) bones.

The **femur**, or thighbone, is the longest, heaviest, and strongest bone in the body. The femur and pelvis form the hip joint at the proximal end. At the distal end, the femur, tibia, and patella form the knee joint. The **patella**, or kneecap, is a triangular bone that lies anterior to the knee joint. The patella is embedded in the tendon of the femoral extensors (quadriceps). It improves knee extension by reducing friction. The **tibia**, or shinbone, is a large bone of the leg that is located directly below the knee. The tibia articulates with the femur at its proximal end, with the fibula and the tarsal bones at its distal end. It is the second largest bone in the human body and is responsible for transmitting the weight of the body from the femur to the foot. The **fibula**, or calf bone, parallels and articulates with the tibia. It does



not articulate with the femur and does not bear weight. The fibula acts as a site for muscle attachment and forms the lateral part of the ankle joint.

The **tarsals** are the seven bones of the ankle. The ankle transmits the weight of the body from the tibia and the fibula to the foot. The **metatarsals** are the five bones of the foot. The phalanges are the 14 bones of the toes. Each toe consists of three phalanges, except for the big toe that has only two ([link](#)). Variations exist in other species; for example, the horse's metacarpals and metatarsals are oriented vertically and do not make contact with the substrate.



This drawing shows the bones of the human foot and ankle, including the metatarsals and the phalanges.

### **Note:**

#### **Evolution Connection**

#### **Evolution of Body Design for Locomotion on Land**

The transition of vertebrates onto land required a number of changes in body design, as movement on land presents a number of challenges for animals that are adapted to movement in water. The buoyancy of water provides a certain amount of lift, and a common form of movement by fish is lateral undulations of the entire body. This back and forth movement

pushes the body against the water, creating forward movement. In most fish, the muscles of paired fins attach to girdles within the body, allowing for some control of locomotion. As certain fish began moving onto land, they retained their lateral undulation form of locomotion (anguilliform). However, instead of pushing against water, their fins or flippers became points of contact with the ground, around which they rotated their bodies. The effect of gravity and the lack of buoyancy on land meant that body weight was suspended on the limbs, leading to increased strengthening and ossification of the limbs. The effect of gravity also required changes to the axial skeleton. Lateral undulations of land animal vertebral columns cause torsional strain. A firmer, more ossified vertebral column became common in terrestrial tetrapods because it reduces strain while providing the strength needed to support the body's weight. In later tetrapods, the vertebrae began allowing for vertical motion rather than lateral flexion. Another change in the axial skeleton was the loss of a direct attachment between the pectoral girdle and the head. This reduced the jarring to the head caused by the impact of the limbs on the ground. The vertebrae of the neck also evolved to allow movement of the head independently of the body.

The appendicular skeleton of land animals is also different from aquatic animals. The shoulders attach to the pectoral girdle through muscles and connective tissue, thus reducing the jarring of the skull. Because of a lateral undulating vertebral column, in early tetrapods, the limbs were splayed out to the side and movement occurred by performing "push-ups." The vertebrae of these animals had to move side-to-side in a similar manner to fish and reptiles. This type of motion requires large muscles to move the limbs toward the midline; it was almost like walking while doing push-ups, and it is not an efficient use of energy. Later tetrapods have their limbs placed under their bodies, so that each stride requires less force to move forward. This resulted in decreased adductor muscle size and an increased range of motion of the scapulae. This also restricts movement primarily to one plane, creating forward motion rather than moving the limbs upward as well as forward. The femur and humerus were also rotated, so that the ends of the limbs and digits were pointed forward, in the direction of motion, rather than out to the side. By placement underneath the body, limbs can swing forward like a pendulum to produce a stride that is more efficient for moving over land.

## Section Summary

The three types of skeleton designs are hydrostatic skeletons, exoskeletons, and endoskeletons. A hydrostatic skeleton is formed by a fluid-filled compartment held under hydrostatic pressure; movement is created by the muscles producing pressure on the fluid. An exoskeleton is a hard external skeleton that protects the outer surface of an organism and enables movement through muscles attached on the inside. An endoskeleton is an internal skeleton composed of hard, mineralized tissue that also enables movement by attachment to muscles. The human skeleton is an endoskeleton that is composed of the axial and appendicular skeleton. The axial skeleton is composed of the bones of the skull, ossicles of the ear, hyoid bone, vertebral column, and ribcage. The skull consists of eight cranial bones and 14 facial bones. Six bones make up the ossicles of the middle ear, while the hyoid bone is located in the neck under the mandible. The vertebral column contains 26 bones, and it surrounds and protects the spinal cord. The thoracic cage consists of the sternum, ribs, thoracic vertebrae, and costal cartilages. The appendicular skeleton is made up of the limbs of the upper and lower limbs. The pectoral girdle is composed of the clavicles and the scapulae. The upper limb contains 30 bones in the arm, the forearm, and the hand. The pelvic girdle attaches the lower limbs to the axial skeleton. The lower limb includes the bones of the thigh, the leg, and the foot.

## Review Questions

### Exercise:

**Problem:** The forearm consists of the:

- a. radius and ulna
- b. radius and humerus
- c. ulna and humerus
- d. humerus and carpus

---

**Solution:**

A

**Exercise:**

**Problem:** The pectoral girdle consists of the:

- a. clavicle and sternum
  - b. sternum and scapula
  - c. clavicle and scapula
  - d. clavicle and coccyx
- 

**Solution:**

C

**Exercise:**

**Problem:**

All of the following are groups of vertebrae except \_\_\_\_\_, which is a curvature.

- a. thoracic
  - b. cervical
  - c. lumbar
  - d. pelvic
- 

**Solution:**

D

**Exercise:**

**Problem:** Which of these is a facial bone?

- a. frontal
- b. occipital
- c. lacrimal

d. temporal

---

**Solution:**

C

**Free Response**

**Exercise:**

**Problem:**

What are the major differences between the male pelvis and female pelvis that permit childbirth in females?

---

**Solution:**

The female pelvis is tilted forward and is wider, lighter, and shallower than the male pelvis. It also has a pubic angle that is broader than the male pelvis.

**Exercise:**

**Problem:**

What are the major differences between the pelvic girdle and the pectoral girdle that allow the pelvic girdle to bear the weight of the body?

---

**Solution:**

The pelvic girdle is securely attached to the body by strong ligaments, unlike the pectoral girdle, which is sparingly attached to the ribcage. The sockets of the pelvic girdle are deep, allowing the femur to be more stable than the pectoral girdle, which has shallow sockets for the scapula. Most tetrapods have 75 percent of their weight on the front legs because the head and neck are so heavy; the advantage of the shoulder joint is more degrees of freedom in movement.

## Glossary

appendicular skeleton

composed of the bones of the upper limbs, which function to grasp and manipulate objects, and the lower limbs, which permit locomotion

articulation

any place where two bones are joined

auditory ossicle

(also, middle ear) transduces sounds from the air into vibrations in the fluid-filled cochlea

axial skeleton

forms the central axis of the body and includes the bones of the skull, the ossicles of the middle ear, the hyoid bone of the throat, the vertebral column, and the thoracic cage (ribcage)

carpus

eight bones that comprise the wrist

clavicle

S-shaped bone that positions the arms laterally

coxal bone

hip bone

cranial bone

one of eight bones that form the cranial cavity that encloses the brain and serves as an attachment site for the muscles of the head and neck

endoskeleton

skeleton of living cells that produce a hard, mineralized tissue located within the soft tissue of organisms

exoskeleton

a secreted cellular product external skeleton that consists of a hard encasement on the surface of an organism

facial bone

one of the 14 bones that form the face; provides cavities for the sense organs (eyes, mouth, and nose) and attachment points for facial muscles

femur

(also, thighbone) longest, heaviest, and strongest bone in the body

fibula

(also, calf bone) parallels and articulates with the tibia

forearm

extends from the elbow to the wrist and consists of two bones: the ulna and the radius

humerus

only bone of the arm

hydrostatic skeleton

skeleton that consists of aqueous fluid held under pressure in a closed body compartment

hyoid bone

lies below the mandible in the front of the neck

intervertebral disc

composed of fibrous cartilage; lies between adjacent vertebrae from the second cervical vertebra to the sacrum

lower limb

consists of the thigh, the leg, and the foot

metacarpus

five bones that comprise the palm

metatarsal

one of the five bones of the foot

patella

(also, kneecap) triangular bone that lies anterior to the knee joint

pectoral girdle

bones that transmit the force generated by the upper limbs to the axial skeleton

phalange

one of the bones of the fingers or toes

pelvic girdle

bones that transmit the force generated by the lower limbs to the axial skeleton

radius

bone located along the lateral (thumb) side of the forearm; articulates with the humerus at the elbow

rib

one of 12 pairs of long, curved bones that attach to the thoracic vertebrae and curve toward the front of the body to form the ribcage

scapula

flat, triangular bone located at the posterior pectoral girdle

skull

bone that supports the structures of the face and protects the brain

sternum

(also, breastbone) long, flat bone located at the front of the chest

tarsal

one of the seven bones of the ankle

thoracic cage

(also, ribcage) skeleton of the chest, which consists of the ribs, thoracic vertebrae, sternum, and costal cartilages



tibia

(also, shinbone) large bone of the leg that is located directly below the knee

ulna

bone located on the medial aspect (pinky-finger side) of the forearm

vertebral column

(also, spine) surrounds and protects the spinal cord, supports the head, and acts as an attachment point for ribs and muscles of the back and neck

## Bone

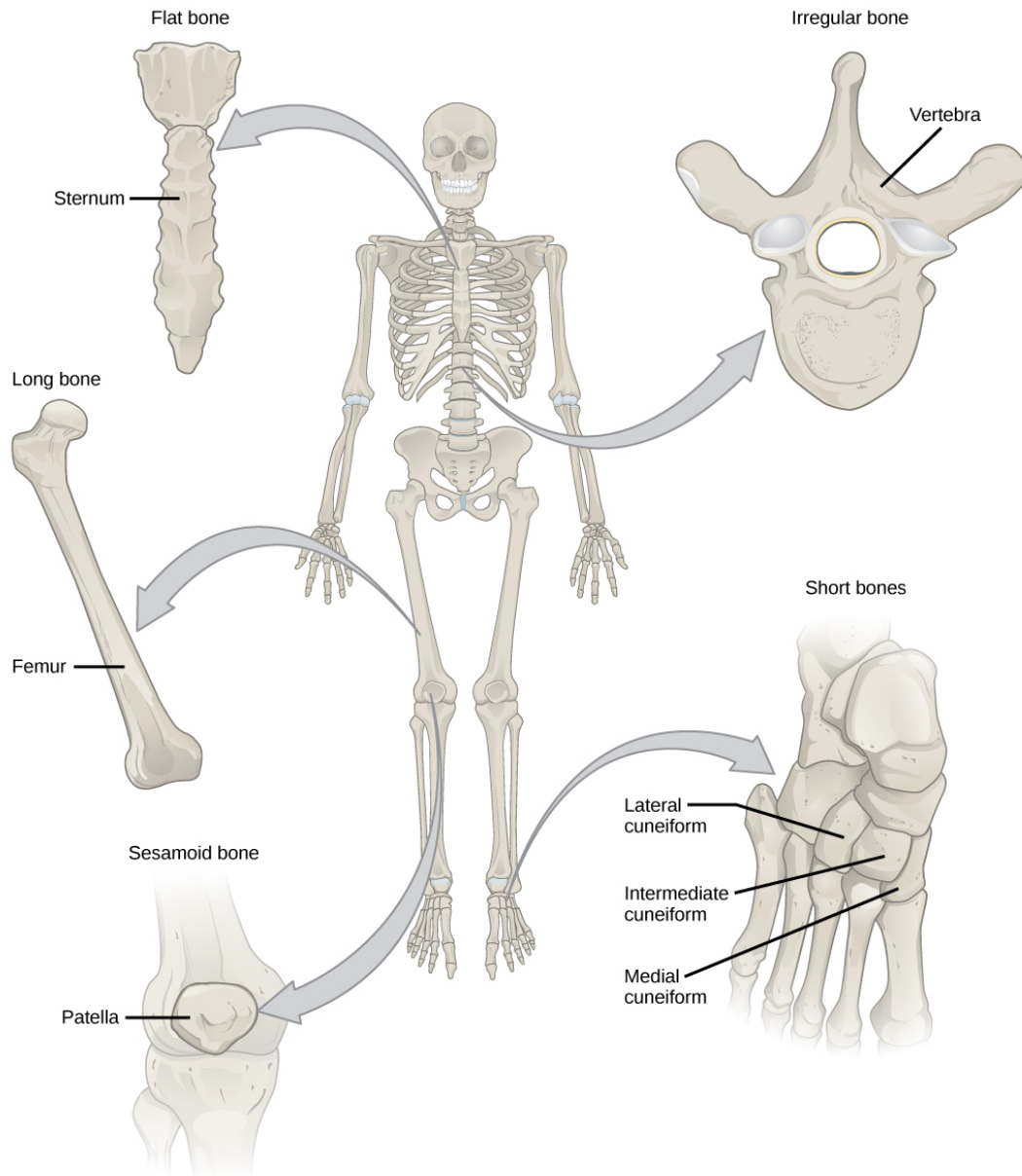
By the end of this section, you will be able to:

- Classify the different types of bones in the skeleton
- Explain the role of the different cell types in bone
- Explain how bone forms during development

**Bone**, or **osseous tissue**, is a connective tissue that constitutes the endoskeleton. It contains specialized cells and a matrix of mineral salts and collagen fibers.

The mineral salts primarily include hydroxyapatite, a mineral formed from calcium phosphate. **Calcification** is the process of deposition of mineral salts on the collagen fiber matrix that crystallizes and hardens the tissue. The process of calcification only occurs in the presence of collagen fibers.

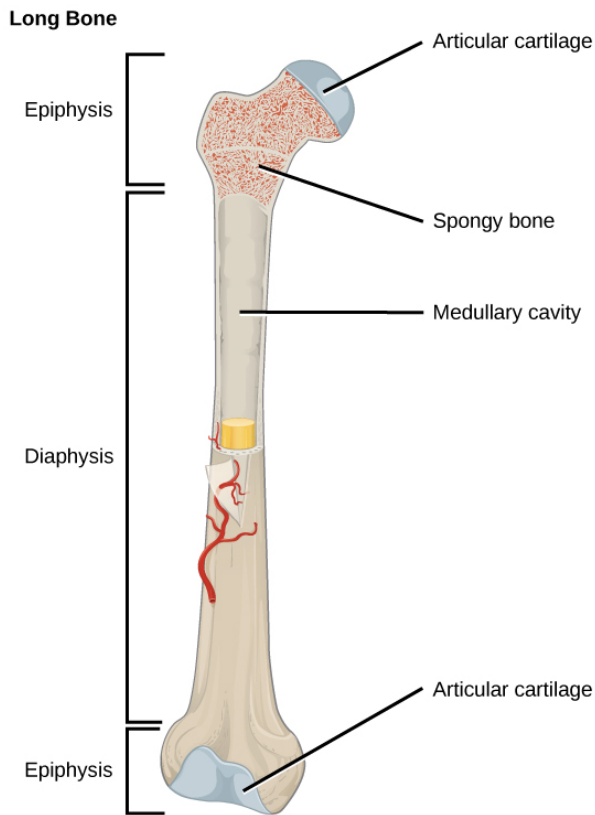
The bones of the human skeleton are classified by their shape: long bones, short bones, flat bones, sutural bones, sesamoid bones, and irregular bones ([link](#)).



Shown are different types of bones: flat, irregular, long, short, and sesamoid.

**Long bones** are longer than they are wide and have a shaft and two ends. The **diaphysis**, or central shaft, contains bone marrow in a marrow cavity. The rounded ends, the **epiphyses**, are covered with articular cartilage and are filled with red bone marrow, which produces blood cells ([\[link\]](#)). Most

of the limb bones are long bones—for example, the femur, tibia, ulna, and radius. Exceptions to this include the patella and the bones of the wrist and ankle.



The long bone is covered by articular cartilage at either end and contains bone marrow (shown in yellow in this illustration) in the marrow cavity.

**Short bones**, or cuboidal bones, are bones that are the same width and length, giving them a cube-like shape. For example, the bones of the wrist (carpals) and ankle (tarsals) are short bones ([link](#)).

**Flat bones** are thin and relatively broad bones that are found where extensive protection of organs is required or where broad surfaces of

muscle attachment are required. Examples of flat bones are the sternum (breast bone), ribs, scapulae (shoulder blades), and the roof of the skull ([link](#)).

**Irregular bones** are bones with complex shapes. These bones may have short, flat, notched, or ridged surfaces. Examples of irregular bones are the vertebrae, hip bones, and several skull bones.

**Sesamoid bones** are small, flat bones and are shaped similarly to a sesame seed. The patellae are sesamoid bones ([link](#)). Sesamoid bones develop inside tendons and may be found near joints at the knees, hands, and feet.



The patella of the knee is an example of a sesamoid bone.

**Sutural bones** are small, flat, irregularly shaped bones. They may be found between the flat bones of the skull. They vary in number, shape, size, and position.

## Bone Tissue

Bones are considered organs because they contain various types of tissue, such as blood, connective tissue, nerves, and bone tissue. Osteocytes, the

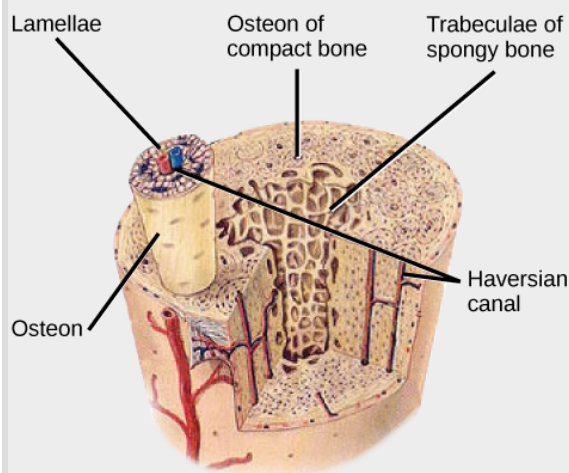
living cells of bone tissue, form the mineral matrix of bones. There are two types of bone tissue: compact and spongy.

## Compact Bone Tissue

**Compact bone** (or cortical bone) forms the hard external layer of all bones and surrounds the medullary cavity, or bone marrow. It provides protection and strength to bones. Compact bone tissue consists of units called osteons or Haversian systems. **Osteons** are cylindrical structures that contain a mineral matrix and living osteocytes connected by canaliculi, which transport blood. They are aligned parallel to the long axis of the bone. Each osteon consists of **lamellae**, which are layers of compact matrix that surround a central canal called the Haversian canal. The **Haversian canal** (osteonic canal) contains the bone's blood vessels and nerve fibers ([\[link\]](#)). Osteons in compact bone tissue are aligned in the same direction along lines of stress and help the bone resist bending or fracturing. Therefore, compact bone tissue is prominent in areas of bone at which stresses are applied in only a few directions.

### Note:

#### Art Connection



Compact bone tissue consists

of osteons that are aligned parallel to the long axis of the bone, and the Haversian canal that contains the bone's blood vessels and nerve fibers. The inner layer of bones consists of spongy bone tissue. The small dark ovals in the osteon represent the living osteocytes. (credit: modification of work by NCI, NIH)

Which of the following statements about bone tissue is false?

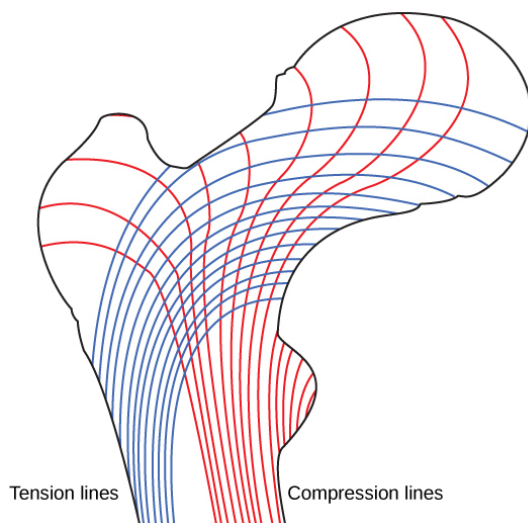
- a. Compact bone tissue is made of cylindrical osteons that are aligned such that they travel the length of the bone.
- b. Haversian canals contain blood vessels only.
- c. Haversian canals contain blood vessels and nerve fibers.
- d. Spongy tissue is found on the interior of the bone, and compact bone tissue is found on the exterior.

## Spongy Bone Tissue

Whereas compact bone tissue forms the outer layer of all bones, **spongy bone** or cancellous bone forms the inner layer of all bones. Spongy bone tissue does not contain osteons that constitute compact bone tissue. Instead, it consists of **trabeculae**, which are lamellae that are arranged as rods or plates. Red bone marrow is found between the trabeculae. Blood vessels within this tissue deliver nutrients to osteocytes and remove waste. The red bone marrow of the femur and the interior of other large bones, such as the ileum, forms blood cells.

Spongy bone reduces the density of bone and allows the ends of long bones to compress as the result of stresses applied to the bone. Spongy bone is prominent in areas of bones that are not heavily stressed or where stresses arrive from many directions. The epiphyses of bones, such as the neck of the femur, are subject to stress from many directions. Imagine laying a heavy framed picture flat on the floor. You could hold up one side of the picture with a toothpick if the toothpick was perpendicular to the floor and the picture. Now drill a hole and stick the toothpick into the wall to hang up the picture. In this case, the function of the toothpick is to transmit the downward pressure of the picture to the wall. The force on the picture is straight down to the floor, but the force on the toothpick is both the picture wire pulling down and the bottom of the hole in the wall pushing up. The toothpick will break off right at the wall.

The neck of the femur is horizontal like the toothpick in the wall. The weight of the body pushes it down near the joint, but the vertical diaphysis of the femur pushes it up at the other end. The neck of the femur must be strong enough to transfer the downward force of the body weight horizontally to the vertical shaft of the femur ([\[link\]](#)).



Trabeculae in spongy bone are arranged such that one side of the bone bears



tension and the other  
withstands compression.

**Note:**

Link to Learning



View [micrographs](#) of musculoskeletal tissues as you review the anatomy.

## Cell Types in Bones

Bone consists of four types of cells: osteoblasts, osteoclasts, osteocytes, and osteoprogenitor cells. **Osteoblasts** are bone cells that are responsible for bone formation. Osteoblasts synthesize and secrete the organic part and inorganic part of the extracellular matrix of bone tissue, and collagen fibers. Osteoblasts become trapped in these secretions and differentiate into less active osteocytes. **Osteoclasts** are large bone cells with up to 50 nuclei. They remove bone structure by releasing lysosomal enzymes and acids that dissolve the bony matrix. These minerals, released from bones into the blood, help regulate calcium concentrations in body fluids. Bone may also be resorbed for remodeling, if the applied stresses have changed.

**Osteocytes** are mature bone cells and are the main cells in bony connective tissue; these cells cannot divide. Osteocytes maintain normal bone structure by recycling the mineral salts in the bony matrix. **Osteoprogenitor cells** are squamous stem cells that divide to produce daughter cells that differentiate

into osteoblasts. Osteoprogenitor cells are important in the repair of fractures.

## **Development of Bone**

**Ossification**, or osteogenesis, is the process of bone formation by osteoblasts. Ossification is distinct from the process of calcification; whereas calcification takes place during the ossification of bones, it can also occur in other tissues. Ossification begins approximately six weeks after fertilization in an embryo. Before this time, the embryonic skeleton consists entirely of fibrous membranes and hyaline cartilage. The development of bone from fibrous membranes is called intramembranous ossification; development from hyaline cartilage is called endochondral ossification. Bone growth continues until approximately age 25. Bones can grow in thickness throughout life, but after age 25, ossification functions primarily in bone remodeling and repair.

### **Intramembranous Ossification**

**Intramembranous ossification** is the process of bone development from fibrous membranes. It is involved in the formation of the flat bones of the skull, the mandible, and the clavicles. Ossification begins as mesenchymal cells form a template of the future bone. They then differentiate into osteoblasts at the ossification center. Osteoblasts secrete the extracellular matrix and deposit calcium, which hardens the matrix. The non-mineralized portion of the bone or osteoid continues to form around blood vessels, forming spongy bone. Connective tissue in the matrix differentiates into red bone marrow in the fetus. The spongy bone is remodeled into a thin layer of compact bone on the surface of the spongy bone.

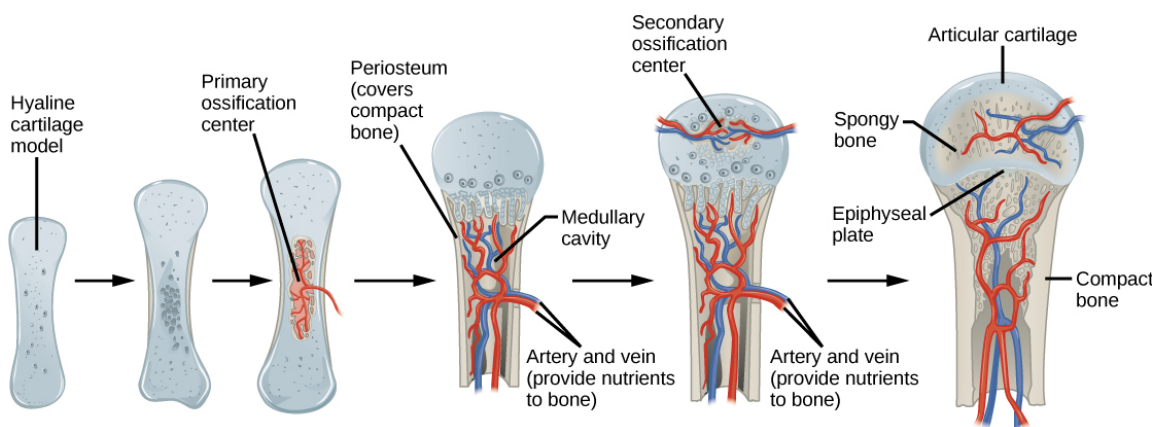
### **Endochondral Ossification**

**Endochondral ossification** is the process of bone development from hyaline cartilage. All of the bones of the body, except for the flat bones of

the skull, mandible, and clavicles, are formed through endochondral ossification.

In long bones, chondrocytes form a template of the hyaline cartilage diaphysis. Responding to complex developmental signals, the matrix begins to calcify. This calcification prevents diffusion of nutrients into the matrix, resulting in chondrocytes dying and the opening up of cavities in the diaphysis cartilage. Blood vessels invade the cavities, and osteoblasts and osteoclasts modify the calcified cartilage matrix into spongy bone. Osteoclasts then break down some of the spongy bone to create a marrow, or medullary, cavity in the center of the diaphysis. Dense, irregular connective tissue forms a sheath (periosteum) around the bones. The periosteum assists in attaching the bone to surrounding tissues, tendons, and ligaments. The bone continues to grow and elongate as the cartilage cells at the epiphyses divide.

In the last stage of prenatal bone development, the centers of the epiphyses begin to calcify. Secondary ossification centers form in the epiphyses as blood vessels and osteoblasts enter these areas and convert hyaline cartilage into spongy bone. Until adolescence, hyaline cartilage persists at the **epiphyseal plate** (growth plate), which is the region between the diaphysis and epiphysis that is responsible for the lengthwise growth of long bones ([\[link\]](#)).



Endochondral ossification is the process of bone development

from hyaline cartilage. The periosteum is the connective tissue on the outside of bone that acts as the interface between bone, blood vessels, tendons, and ligaments.

## **Growth of Bone**

Long bones continue to lengthen, potentially until adolescence, through the addition of bone tissue at the epiphyseal plate. They also increase in width through appositional growth.

### **Lengthening of Long Bones**

Chondrocytes on the epiphyseal side of the epiphyseal plate divide; one cell remains undifferentiated near the epiphysis, and one cell moves toward the diaphysis. The cells, which are pushed from the epiphysis, mature and are destroyed by calcification. This process replaces cartilage with bone on the diaphyseal side of the plate, resulting in a lengthening of the bone.

Long bones stop growing at around the age of 18 in females and the age of 21 in males in a process called epiphyseal plate closure. During this process, cartilage cells stop dividing and all of the cartilage is replaced by bone. The epiphyseal plate fades, leaving a structure called the epiphyseal line or epiphyseal remnant, and the epiphysis and diaphysis fuse.

### **Thickening of Long Bones**

**Appositional growth** is the increase in the diameter of bones by the addition of bony tissue at the surface of bones. Osteoblasts at the bone surface secrete bone matrix, and osteoclasts on the inner surface break down bone. The osteoblasts differentiate into osteocytes. A balance between these two processes allows the bone to thicken without becoming too heavy.

## Bone Remodeling and Repair

Bone renewal continues after birth into adulthood. **Bone remodeling** is the replacement of old bone tissue by new bone tissue. It involves the processes of bone deposition by osteoblasts and bone resorption by osteoclasts. Normal bone growth requires vitamins D, C, and A, plus minerals such as calcium, phosphorous, and magnesium. Hormones such as parathyroid hormone, growth hormone, and calcitonin are also required for proper bone growth and maintenance.

Bone turnover rates are quite high, with five to seven percent of bone mass being recycled every week. Differences in turnover rate exist in different areas of the skeleton and in different areas of a bone. For example, the bone in the head of the femur may be fully replaced every six months, whereas the bone along the shaft is altered much more slowly.

Bone remodeling allows bones to adapt to stresses by becoming thicker and stronger when subjected to stress. Bones that are not subject to normal stress, for example when a limb is in a cast, will begin to lose mass. A fractured or broken bone undergoes repair through four stages:

1. Blood vessels in the broken bone tear and hemorrhage, resulting in the formation of clotted blood, or a hematoma, at the site of the break. The severed blood vessels at the broken ends of the bone are sealed by the clotting process, and bone cells that are deprived of nutrients begin to die.
2. Within days of the fracture, capillaries grow into the hematoma, and phagocytic cells begin to clear away the dead cells. Though fragments of the blood clot may remain, fibroblasts and osteoblasts enter the area and begin to reform bone. Fibroblasts produce collagen fibers that connect the broken bone ends, and osteoblasts start to form spongy bone. The repair tissue between the broken bone ends is called the fibrocartilaginous callus, as it is composed of both hyaline and fibrocartilage ([link](#)). Some bone spicules may also appear at this point.
3. The fibrocartilaginous callus is converted into a bony callus of spongy bone. It takes about two months for the broken bone ends to be firmly

joined together after the fracture. This is similar to the endochondral formation of bone, as cartilage becomes ossified; osteoblasts, osteoclasts, and bone matrix are present.

4. The bony callus is then remodelled by osteoclasts and osteoblasts, with excess material on the exterior of the bone and within the medullary cavity being removed. Compact bone is added to create bone tissue that is similar to the original, unbroken bone. This remodeling can take many months, and the bone may remain uneven for years.



After this bone is set, a callus will knit the two ends together.

(credit: Bill Rhodes)

**Note:**

Scientific Method Connection  
**Decalcification of Bones**

**Question:** What effect does the removal of calcium and collagen have on bone structure?

**Background:** Conduct a literature search on the role of calcium and collagen in maintaining bone structure. Conduct a literature search on diseases in which bone structure is compromised.

**Hypothesis:** Develop a hypothesis that states predictions of the flexibility, strength, and mass of bones that have had the calcium and collagen components removed. Develop a hypothesis regarding the attempt to add calcium back to decalcified bones.

**Test the hypothesis:** Test the prediction by removing calcium from chicken bones by placing them in a jar of vinegar for seven days. Test the hypothesis regarding adding calcium back to decalcified bone by placing the decalcified chicken bones into a jar of water with calcium supplements added. Test the prediction by denaturing the collagen from the bones by baking them at 250°C for three hours.

**Analyze the data:** Create a table showing the changes in bone flexibility, strength, and mass in the three different environments.

**Report the results:** Under which conditions was the bone most flexible? Under which conditions was the bone the strongest?

**Draw a conclusion:** Did the results support or refute the hypothesis? How do the results observed in this experiment correspond to diseases that destroy bone tissue?

## Section Summary

Bone, or osseous tissue, is connective tissue that includes specialized cells, mineral salts, and collagen fibers. The human skeleton can be divided into long bones, short bones, flat bones, and irregular bones. Compact bone tissue is composed of osteons and forms the external layer of all bones. Spongy bone tissue is composed of trabeculae and forms the inner part of all bones. Four types of cells compose bony tissue: osteocytes, osteoclasts, osteoprogenitor cells, and osteoblasts. Ossification is the process of bone formation by osteoblasts. Intramembranous ossification is the process of bone development from fibrous membranes. Endochondral ossification is the process of bone development from hyaline cartilage. Long bones

lengthen as chondrocytes divide and secrete hyaline cartilage. Osteoblasts replace cartilage with bone. Appositional growth is the increase in the diameter of bones by the addition of bone tissue at the surface of bones. Bone remodeling involves the processes of bone deposition by osteoblasts and bone resorption by osteoclasts. Bone repair occurs in four stages and can take several months.

## Art Exercise

### Exercise:

#### Problem:

[\[link\]](#) Which of the following statements about bone tissue is false?

- a. Compact bone tissue is made of cylindrical osteons that are aligned such that they travel the length of the bone.
- b. Haversian canals contain blood vessels only.
- c. Haversian canals contain blood vessels and nerve fibers.
- d. Spongy tissue is found on the interior of the bone, and compact bone tissue is found on the exterior.

---

#### Solution:

[\[link\]](#)B

## Review Questions

### Exercise:

#### Problem: The Haversian canal:

- a. is arranged as rods or plates
- b. contains the bone's blood vessels and nerve fibers
- c. is responsible for the lengthwise growth of long bones
- d. synthesizes and secretes matrix



---

**Solution:**

B

**Exercise:**

**Problem:**The epiphyseal plate:

- a. is arranged as rods or plates
- b. contains the bone's blood vessels and nerve fibers
- c. is responsible for the lengthwise growth of long bones
- d. synthesizes and secretes bone matrix

---

**Solution:**

C

**Exercise:**

**Problem:**The cells responsible for bone resorption are \_\_\_\_\_.

- a. osteoclasts
- b. osteoblasts
- c. fibroblasts
- d. osteocytes

---

**Solution:**

A

**Exercise:**

**Problem:**Compact bone is composed of \_\_\_\_\_.

- a. trabeculae

- b. compacted collagen
- c. osteons
- d. calcium phosphate only

---

**Solution:**

C

**Free Response**

**Exercise:**

**Problem:**

What are the major differences between spongy bone and compact bone?

---

**Solution:**

Compact bone tissue forms the hard external layer of all bones and consists of osteons. Compact bone tissue is prominent in areas of bone at which stresses are applied in only a few directions. Spongy bone tissue forms the inner layer of all bones and consists of trabeculae. Spongy bone is prominent in areas of bones that are not heavily stressed or at which stresses arrive from many directions.

**Exercise:**

**Problem:**

What are the roles of osteoblasts, osteocytes, and osteoclasts?

---

**Solution:**

Osteocytes function in the exchange of nutrients and wastes with the blood. They also maintain normal bone structure by recycling the mineral salts in the bony matrix. Osteoclasts remove bone tissue by

releasing lysosomal enzymes and acids that dissolve the bony matrix.  
Osteoblasts are bone cells that are responsible for bone formation.

## **Glossary**

appositional growth

increase in the diameter of bones by the addition of bone tissue at the surface of bones

bone

(also, osseous tissue) connective tissue that constitutes the endoskeleton

bone remodeling

replacement of old bone tissue by new bone tissue

calcification

process of deposition of mineral salts in the collagen fiber matrix that crystallizes and hardens the tissue

compact bone

forms the hard external layer of all bones

diaphysis

central shaft of bone, contains bone marrow in a marrow cavity

endochondral ossification

process of bone development from hyaline cartilage

epiphyseal plate

region between the diaphysis and epiphysis that is responsible for the lengthwise growth of long bones

epiphysis

rounded end of bone, covered with articular cartilage and filled with red bone marrow, which produces blood cells

flat bone

thin and relatively broad bone found where extensive protection of organs is required or where broad surfaces of muscle attachment are required

Haversian canal

contains the bone's blood vessels and nerve fibers

intramembranous ossification

process of bone development from fibrous membranes

irregular bone

bone with complex shapes; examples include vertebrae and hip bones

lamella

layer of compact tissue that surrounds a central canal called the Haversian canal

long bone

bone that is longer than wide, and has a shaft and two ends

osteoblast

bone cell responsible for bone formation

osteoclast

large bone cells with up to 50 nuclei, responsible for bone remodeling

osteocyte

mature bone cells and the main cell in bone tissue

osseous tissue

connective tissue that constitutes the endoskeleton

ossification

(also, osteogenesis) process of bone formation by osteoblasts

osteon

cylindrical structure aligned parallel to the long axis of the bone

resorption

process by which osteoclasts release minerals stored in bones

sesamoid bone

small, flat bone shaped like a sesame seed; develops inside tendons

short bone

bone that has the same width and length, giving it a cube-like shape

spongy bone tissue

forms the inner layer of all bones

suture bone

small, flat, irregularly shaped bone that forms between the flat bones of the cranium

trabeculae

lamellae that are arranged as rods or plates

## Joints and Skeletal Movement

By the end of this section, you will be able to:

- Classify the different types of joints on the basis of structure
- Explain the role of joints in skeletal movement

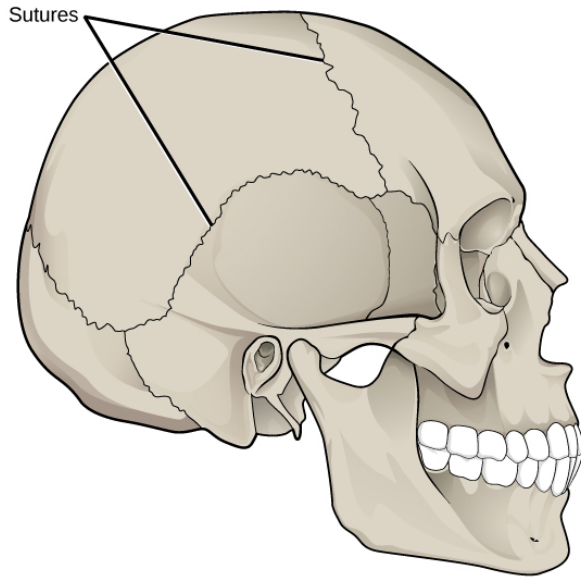
The point at which two or more bones meet is called a **joint**, or **articulation**. Joints are responsible for movement, such as the movement of limbs, and stability, such as the stability found in the bones of the skull.

### Classification of Joints on the Basis of Structure

There are two ways to classify joints: on the basis of their structure or on the basis of their function. The structural classification divides joints into bony, fibrous, cartilaginous, and synovial joints depending on the material composing the joint and the presence or absence of a cavity in the joint.

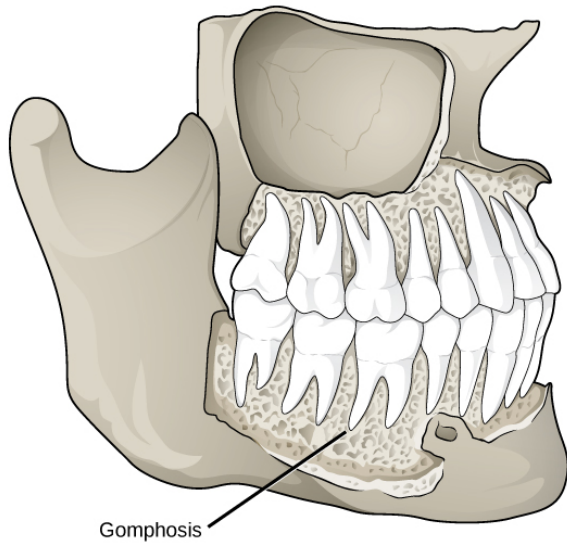
#### Fibrous Joints

The bones of **fibrous joints** are held together by fibrous connective tissue. There is no cavity, or space, present between the bones and so most fibrous joints do not move at all, or are only capable of minor movements. There are three types of fibrous joints: sutures, syndesmoses, and gomphoses. **Sutures** are found only in the skull and possess short fibers of connective tissue that hold the skull bones tightly in place ([link](#)).



Sutures are fibrous joints found only in the skull.

**Syndesmoses** are joints in which the bones are connected by a band of connective tissue, allowing for more movement than in a suture. An example of a syndesmosis is the joint of the tibia and fibula in the ankle. The amount of movement in these types of joints is determined by the length of the connective tissue fibers. **Gomphoses** occur between teeth and their sockets; the term refers to the way the tooth fits into the socket like a peg ([link](#)). The tooth is connected to the socket by a connective tissue referred to as the periodontal ligament.



Gomphoses are fibrous joints between the teeth and their sockets. (credit: modification of work by Gray's Anatomy)

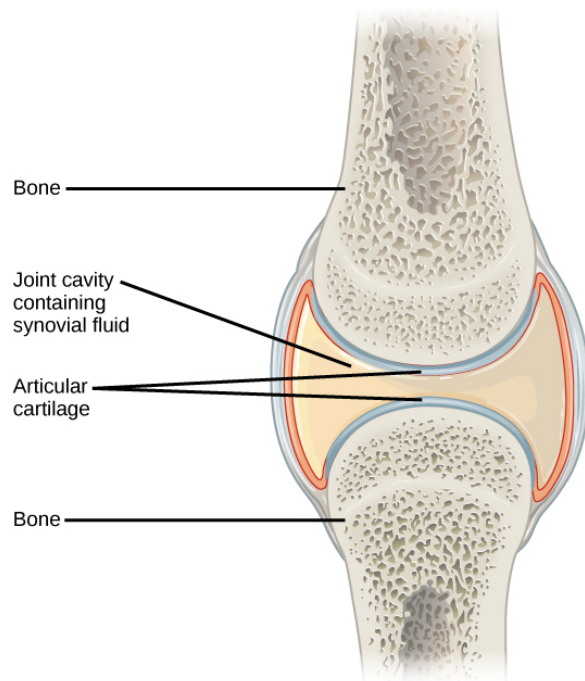
## Cartilaginous Joints

**Cartilaginous joints** are joints in which the bones are connected by cartilage. There are two types of cartilaginous joints: synchondroses and symphyses. In a **synchondrosis**, the bones are joined by hyaline cartilage. Synchondroses are found in the epiphyseal plates of growing bones in children. In **symphyses**, hyaline cartilage covers the end of the bone but the connection between bones occurs through fibrocartilage. Symphyses are found at the joints between vertebrae. Either type of cartilaginous joint allows for very little movement.

## Synovial Joints



**Synovial joints** are the only joints that have a space between the adjoining bones ([\[link\]](#)). This space is referred to as the synovial (or joint) cavity and is filled with synovial fluid. Synovial fluid lubricates the joint, reducing friction between the bones and allowing for greater movement. The ends of the bones are covered with articular cartilage, a hyaline cartilage, and the entire joint is surrounded by an articular capsule composed of connective tissue that allows movement of the joint while resisting dislocation. Articular capsules may also possess ligaments that hold the bones together. Synovial joints are capable of the greatest movement of the three structural joint types; however, the more mobile a joint, the weaker the joint. Knees, elbows, and shoulders are examples of synovial joints.



Synovial joints are the only joints that have a space or “synovial cavity” in the joint.

## Classification of Joints on the Basis of Function

The functional classification divides joints into three categories: synarthroses, amphiarthroses, and diarthroses. A **synarthrosis** is a joint that is immovable. This includes sutures, gomphoses, and synchondroses.

**Amphiarthroses** are joints that allow slight movement, including syndesmoses and symphyses. **Diarthroses** are joints that allow for free movement of the joint, as in synovial joints.

## Movement at Synovial Joints

The wide range of movement allowed by synovial joints produces different types of movements. The movement of synovial joints can be classified as one of four different types: gliding, angular, rotational, or special movement.

### Gliding Movement

**Gliding movements** occur as relatively flat bone surfaces move past each other. Gliding movements produce very little rotation or angular movement of the bones. The joints of the carpal and tarsal bones are examples of joints that produce gliding movements.

### Angular Movement

**Angular movements** are produced when the angle between the bones of a joint changes. There are several different types of angular movements, including flexion, extension, hyperextension, abduction, adduction, and circumduction. **Flexion**, or bending, occurs when the angle between the bones decreases. Moving the forearm upward at the elbow or moving the wrist to move the hand toward the forearm are examples of flexion.

**Extension** is the opposite of flexion in that the angle between the bones of a joint increases. Straightening a limb after flexion is an example of extension. Extension past the regular anatomical position is referred to as **hyperextension**. This includes moving the neck back to look upward, or bending the wrist so that the hand moves away from the forearm.

**Abduction** occurs when a bone moves away from the midline of the body. Examples of abduction are moving the arms or legs laterally to lift them straight out to the side. **Adduction** is the movement of a bone toward the midline of the body. Movement of the limbs inward after abduction is an example of adduction. **Circumduction** is the movement of a limb in a circular motion, as in moving the arm in a circular motion.

## Rotational Movement

**Rotational movement** is the movement of a bone as it rotates around its longitudinal axis. Rotation can be toward the midline of the body, which is referred to as **medial rotation**, or away from the midline of the body, which is referred to as **lateral rotation**. Movement of the head from side to side is an example of rotation.

## Special Movements

Some movements that cannot be classified as gliding, angular, or rotational are called special movements. **Inversion** involves the soles of the feet moving inward, toward the midline of the body. **Eversion** is the opposite of inversion, movement of the sole of the foot outward, away from the midline of the body. **Protraction** is the anterior movement of a bone in the horizontal plane. **Retraction** occurs as a joint moves back into position after protraction. Protraction and retraction can be seen in the movement of the mandible as the jaw is thrust outwards and then back inwards.

**Elevation** is the movement of a bone upward, such as when the shoulders are shrugged, lifting the scapulae. **Depression** is the opposite of elevation—movement downward of a bone, such as after the shoulders are shrugged and the scapulae return to their normal position from an elevated position.

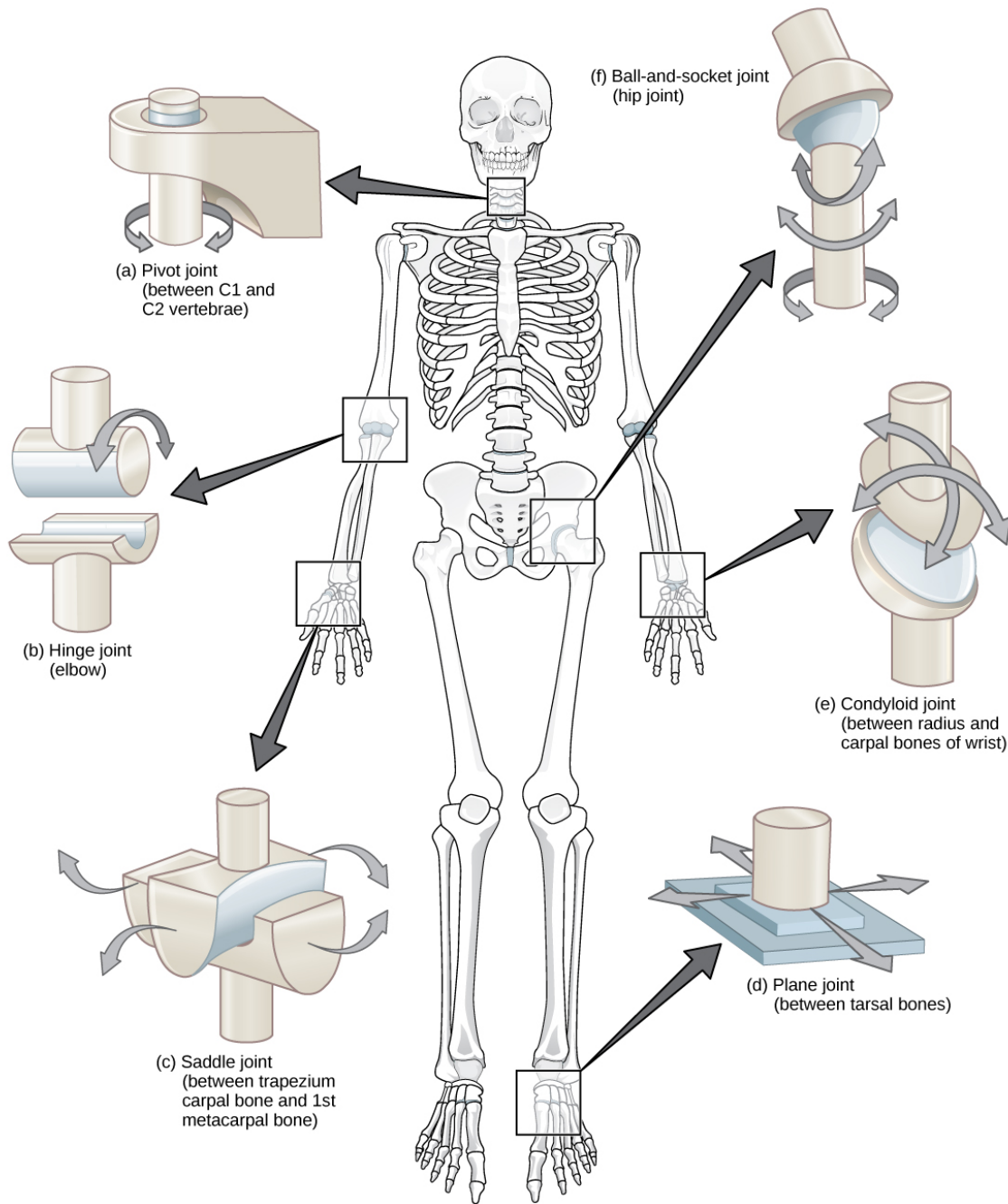
**Dorsiflexion** is a bending at the ankle such that the toes are lifted toward the knee. **Plantar flexion** is a bending at the ankle when the heel is lifted, such as when standing on the toes. **Supination** is the movement of the radius and ulna bones of the forearm so that the palm faces forward.

**Pronation** is the opposite movement, in which the palm faces backward.

**Opposition** is the movement of the thumb toward the fingers of the same hand, making it possible to grasp and hold objects.

## **Types of Synovial Joints**

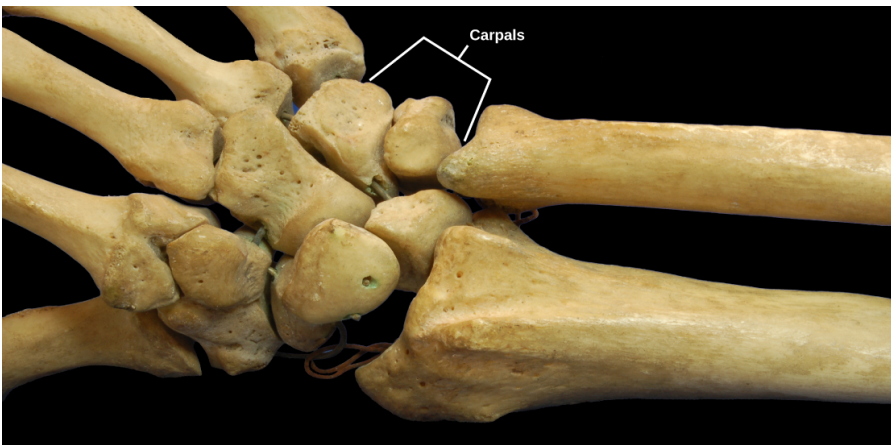
Synovial joints are further classified into six different categories on the basis of the shape and structure of the joint. The shape of the joint affects the type of movement permitted by the joint ([\[link\]](#)). These joints can be described as planar, hinge, pivot, condyloid, saddle, or ball-and-socket joints.



Different types of joints allow different types of movement. Planar, hinge, pivot, condyloid, saddle, and ball-and-socket are all types of synovial joints.

## Planar Joints

**Planar joints** have bones with articulating surfaces that are flat or slightly curved faces. These joints allow for gliding movements, and so the joints are sometimes referred to as gliding joints. The range of motion is limited in these joints and does not involve rotation. Planar joints are found in the carpal bones in the hand and the tarsal bones of the foot, as well as between vertebrae ([link](#)).



The joints of the carpal bones in the wrist are examples of planar joints. (credit: modification of work by Brian C. Goss)

## Hinge Joints

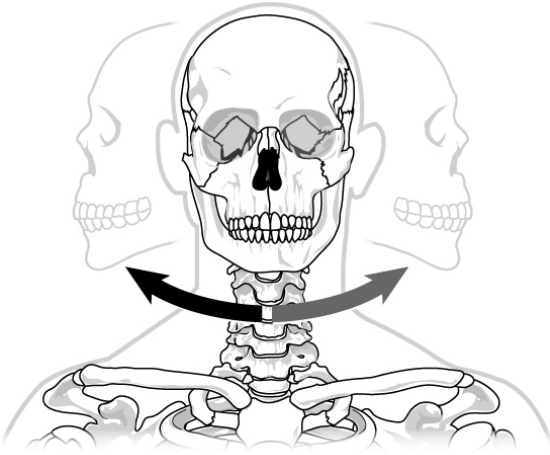
In **hinge joints**, the slightly rounded end of one bone fits into the slightly hollow end of the other bone. In this way, one bone moves while the other remains stationary, like the hinge of a door. The elbow is an example of a hinge joint. The knee is sometimes classified as a modified hinge joint ([link](#)).



The elbow joint, where the radius articulates with the humerus, is an example of a hinge joint. (credit: modification of work by Brian C. Goss)

## Pivot Joints

**Pivot joints** consist of the rounded end of one bone fitting into a ring formed by the other bone. This structure allows rotational movement, as the rounded bone moves around its own axis. An example of a pivot joint is the joint of the first and second vertebrae of the neck that allows the head to move back and forth ([link](#)). The joint of the wrist that allows the palm of the hand to be turned up and down is also a pivot joint.

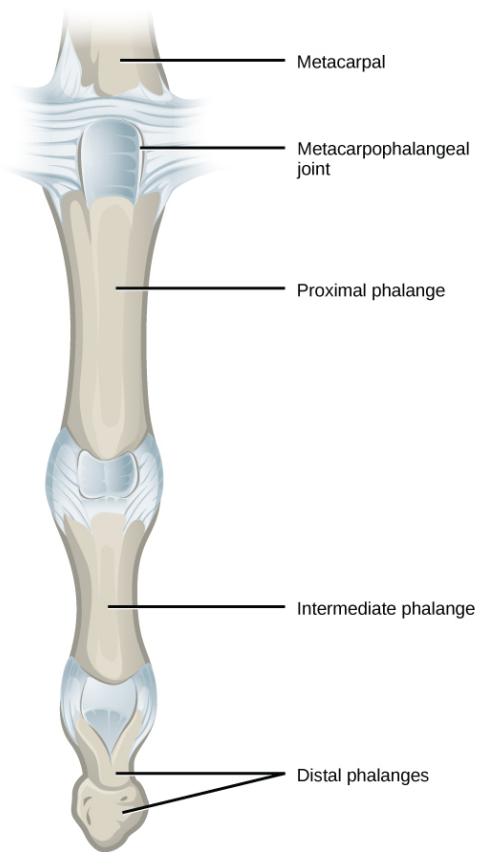


The joint in the neck that allows the head to move back and forth is an example of a pivot joint.

## Condyloid Joints

**Condyloid joints** consist of an oval-shaped end of one bone fitting into a similarly oval-shaped hollow of another bone ([link](#)). This is also sometimes called an ellipsoidal joint. This type of joint allows angular movement along two axes, as seen in the joints of the wrist and fingers, which can move both side to side and up and down.



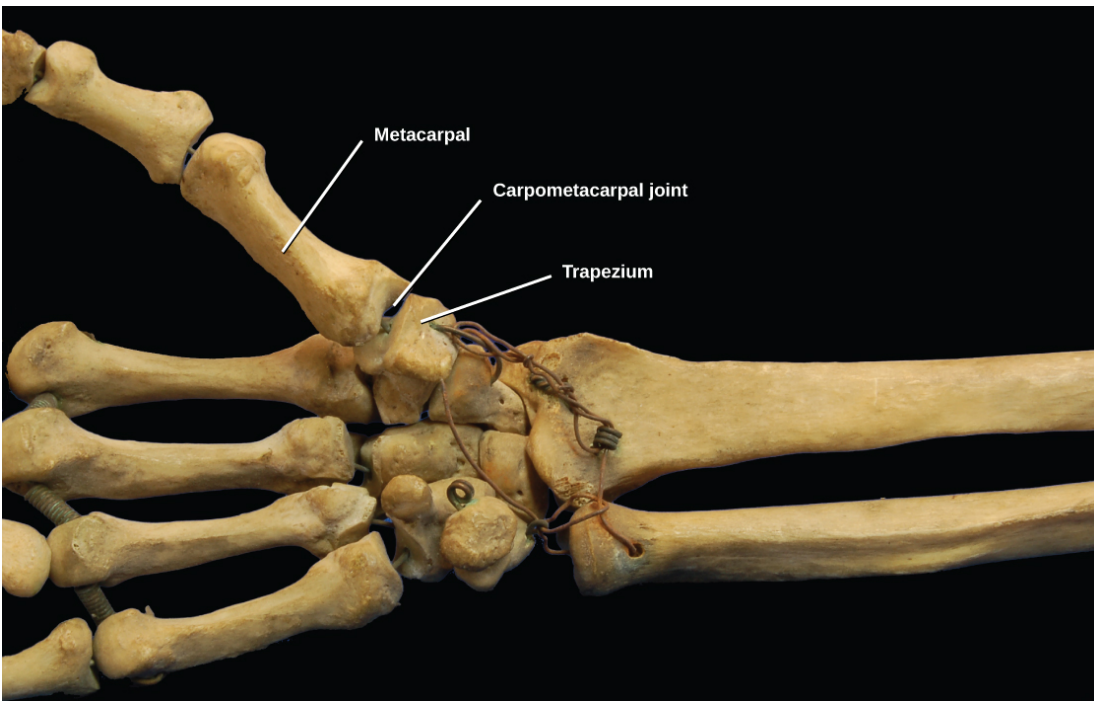


The metacarpophalangeal joints in the finger are examples of condyloid joints. (credit: modification of work by Gray's Anatomy)

## Saddle Joints

**Saddle joints** are so named because the ends of each bone resemble a saddle, with concave and convex portions that fit together. Saddle joints allow angular movements similar to condyloid joints but with a greater range of motion. An example of a saddle joint is the thumb joint, which can

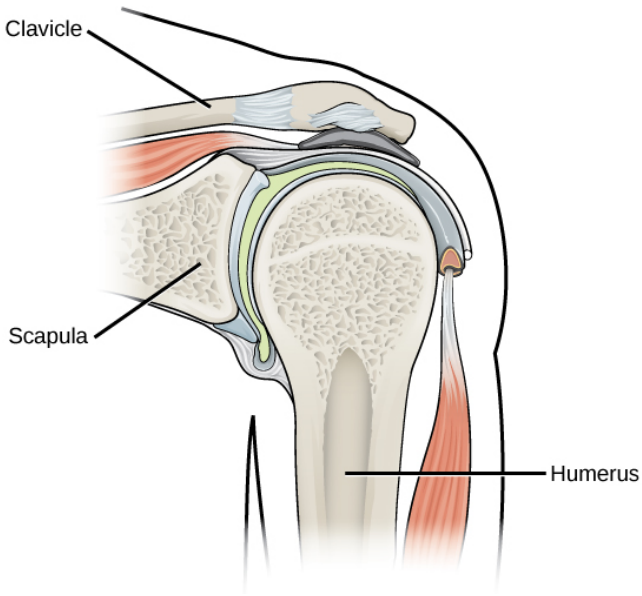
move back and forth and up and down, but more freely than the wrist or fingers ([\[link\]](#)).



The carpometacarpal joints in the thumb are examples of saddle joints. (credit: modification of work by Brian C. Goss)

## Ball-and-Socket Joints

**Ball-and-socket joints** possess a rounded, ball-like end of one bone fitting into a cuplike socket of another bone. This organization allows the greatest range of motion, as all movement types are possible in all directions. Examples of ball-and-socket joints are the shoulder and hip joints ([\[link\]](#)).



The shoulder joint is an example of a ball-and-socket joint.

**Note:**

Link to Learning



Watch this animation showing the six types of synovial joints.

[https://www.openstaxcollege.org/l/synovial\\_joints](https://www.openstaxcollege.org/l/synovial_joints)

**Note:**

Career Connection

## **Rheumatologist**

Rheumatologists are medical doctors who specialize in the diagnosis and treatment of disorders of the joints, muscles, and bones. They diagnose and treat diseases such as arthritis, musculoskeletal disorders, osteoporosis, and autoimmune diseases such as ankylosing spondylitis and rheumatoid arthritis.

Rheumatoid arthritis (RA) is an inflammatory disorder that primarily affects the synovial joints of the hands, feet, and cervical spine. Affected joints become swollen, stiff, and painful. Although it is known that RA is an autoimmune disease in which the body's immune system mistakenly attacks healthy tissue, the cause of RA remains unknown. Immune cells from the blood enter joints and the synovium causing cartilage breakdown, swelling, and inflammation of the joint lining. Breakdown of cartilage causes bones to rub against each other causing pain. RA is more common in women than men and the age of onset is usually 40–50 years of age.

Rheumatologists can diagnose RA on the basis of symptoms such as joint inflammation and pain, X-ray and MRI imaging, and blood tests.

Arthrography is a type of medical imaging of joints that uses a contrast agent, such as a dye, that is opaque to X-rays. This allows the soft tissue structures of joints—such as cartilage, tendons, and ligaments—to be visualized. An arthrogram differs from a regular X-ray by showing the surface of soft tissues lining the joint in addition to joint bones. An arthrogram allows early degenerative changes in joint cartilage to be detected before bones become affected.

There is currently no cure for RA; however, rheumatologists have a number of treatment options available. Early stages can be treated with rest of the affected joints by using a cane or by using joint splints that minimize inflammation. When inflammation has decreased, exercise can be used to strengthen the muscles that surround the joint and to maintain joint flexibility. If joint damage is more extensive, medications can be used to relieve pain and decrease inflammation. Anti-inflammatory drugs such as aspirin, topical pain relievers, and corticosteroid injections may be used. Surgery may be required in cases in which joint damage is severe.

## **Section Summary**

The structural classification of joints divides them into bony, fibrous, cartilaginous, and synovial joints. The bones of fibrous joints are held together by fibrous connective tissue; the three types of fibrous joints are sutures, syndesmoses, and gomphoses. Cartilaginous joints are joints in which the bones are connected by cartilage; the two types of cartilaginous joints are synchondroses and symphyses. Synovial joints are joints that have a space between the adjoining bones. The functional classification divides joints into three categories: synarthroses, amphiarthroses, and diarthroses. The movement of synovial joints can be classified as one of four different types: gliding, angular, rotational, or special movement. Gliding movements occur as relatively flat bone surfaces move past each other. Angular movements are produced when the angle between the bones of a joint changes. Rotational movement is the movement of a bone as it rotates around its own longitudinal axis. Special movements include inversion, eversion, protraction, retraction, elevation, depression, dorsiflexion, plantar flexion, supination, pronation, and opposition. Synovial joints are also classified into six different categories on the basis of the shape and structure of the joint: planar, hinge, pivot, condyloid, saddle, and ball-and-socket.

## Review Questions

### Exercise:

**Problem:** Synchondroses and symphyses are:

- a. synovial joints
- b. cartilaginous joints
- c. fibrous joints
- d. condyloid joints

---

### Solution:

B

### Exercise:

**Problem:**

The movement of bone away from the midline of the body is called \_\_\_\_\_.

- a. circumduction
- b. extension
- c. adduction
- d. abduction

---

**Solution:**

D

**Exercise:**

**Problem:**

Which of the following is not a characteristic of the synovial fluid?

- a. lubrication
- b. shock absorption
- c. regulation of water balance in the joint
- d. protection of articular cartilage

---

**Solution:**

C

**Exercise:**

**Problem:** The elbow is an example of which type of joint?

- a. hinge
- b. pivot
- c. saddle
- d. gliding

---

**Solution:**

A

**Free Response****Exercise:****Problem:**

What movements occur at the hip joint and knees as you bend down to touch your toes?

---

**Solution:**

The hip joint is flexed and the knees are extended.

**Exercise:****Problem:**

What movement(s) occur(s) at the scapulae when you shrug your shoulders?

---

**Solution:**

Elevation is the movement of a bone upward, such as when the shoulders are shrugged, lifting the scapulae. Depression is the downward movement of a bone, such as after the shoulders are shrugged and the scapulae return to their normal position from an elevated position.

**Glossary**

abduction

when a bone moves away from the midline of the body

adduction

movement of the limbs inward after abduction

amphiarthrosis

joint that allows slight movement; includes syndesmoses and symphyses

angular movement

produced when the angle between the bones of a joint changes

ball-and-socket joint

joint with a rounded, ball-like end of one bone fitting into a cuplike socket of another bone

cartilaginous joint

joint in which the bones are connected by cartilage

circumduction

movement of a limb in a circular motion.

condyloid joint

oval-shaped end of one bone fitting into a similarly oval-shaped hollow of another bone

depression

movement downward of a bone, such as after the shoulders are shrugged and the scapulae return to their normal position from an elevated position; opposite of elevation

diarthrosis

joint that allows for free movement of the joint; found in synovial joints

dorsiflexion

bending at the ankle such that the toes are lifted toward the knee

elevation



movement of a bone upward, such as when the shoulders are shrugged, lifting the scapulae

eversion

movement of the sole of the foot outward, away from the midline of the body; opposite of inversion

extension

movement in which the angle between the bones of a joint increases; opposite of flexion

fibrous joint

joint held together by fibrous connective tissue

flexion

movement in which the angle between the bones decreases; opposite of extension

gliding movement

when relatively flat bone surfaces move past each other

gomphosis

the joint in which the tooth fits into the socket like a peg

hinge joint

slightly rounded end of one bone fits into the slightly hollow end of the other bone

hyperextension

extension past the regular anatomical position

inversion

soles of the feet moving inward, toward the midline of the body

joint

point at which two or more bones meet

lateral rotation

rotation away from the midline of the body

medial rotation

rotation toward the midline of the body

opposition

movement of the thumb toward the fingers of the same hand, making it possible to grasp and hold objects

plantar flexion

bending at the ankle such that the heel is lifted, such as when standing on the toes

planar joint

joint with bones whose articulating surfaces are flat

pivot joint

joint with the rounded end of one bone fitting into a ring formed by the other bone

pronation

movement in which the palm faces backward

protraction

anterior movement of a bone in the horizontal plane

retraction

movement in which a joint moves back into position after protraction

rotational movement

movement of a bone as it rotates around its own longitudinal axis

saddle joint

joint with concave and convex portions that fit together; named because the ends of each bone resemble a saddle

supination

movement of the radius and ulna bones of the forearm so that the palm faces forward

suture

short fiber of connective tissue that holds the skull bones tightly in place; found only in the skull

synarthrosis

joint that is immovable

symphysis

hyaline cartilage covers the end of the bone, but the connection between bones occurs through fibrocartilage; symphyses are found at the joints between vertebrae

synchondrosis

bones joined by hyaline cartilage; synchondroses are found in the epiphyseal plates of growing bones in children

syndesmosis

joint in which the bones are connected by a band of connective tissue, allowing for more movement than in a suture

synovial joint

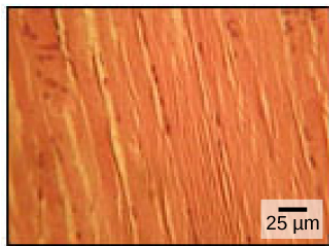
only joint that has a space between the adjoining bones

## Muscle Contraction and Locomotion

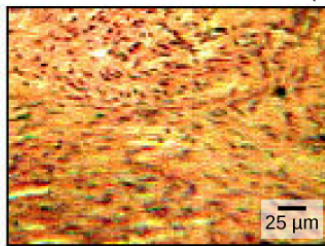
By the end of this section, you will be able to:

- Classify the different types of muscle tissue
- Explain the role of muscles in locomotion

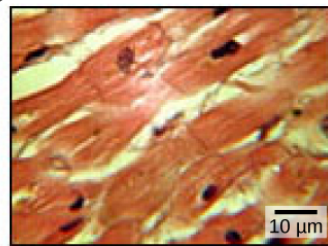
Muscle cells are specialized for contraction. Muscles allow for motions such as walking, and they also facilitate bodily processes such as respiration and digestion. The body contains three types of muscle tissue: skeletal muscle, cardiac muscle, and smooth muscle ([\[link\]](#)).



Skeletal muscle



Smooth muscle



Cardiac muscle

The body contains three types of muscle tissue: skeletal muscle, smooth muscle, and cardiac muscle, visualized here using light microscopy. Smooth muscle cells are short, tapered at each end, and have only one plump nucleus in each. Cardiac muscle cells are branched and striated, but short. The cytoplasm may branch, and they have one nucleus in the center of the cell. (credit: modification of work by NCI, NIH; scale-bar data from Matt Russell)

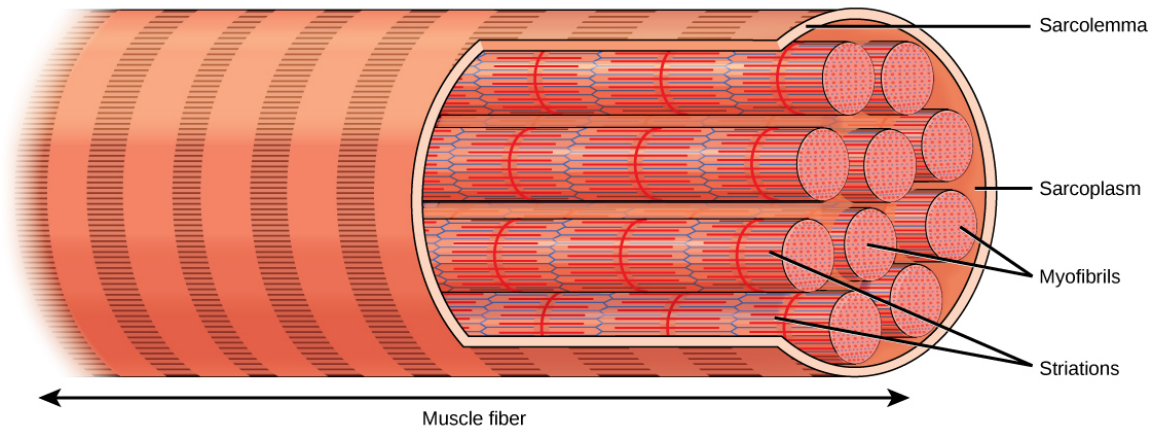
**Skeletal muscle tissue** forms skeletal muscles, which attach to bones or skin and control locomotion and any movement that can be consciously controlled. Because it can be controlled by thought, skeletal muscle is also called voluntary muscle. Skeletal muscles are long and cylindrical in appearance; when viewed under a microscope, skeletal muscle tissue has a striped or striated appearance. The striations are caused by the regular arrangement of contractile proteins (actin and myosin). **Actin** is a globular contractile protein that interacts with **myosin** for muscle contraction. Skeletal muscle also has multiple nuclei present in a single cell.

**Smooth muscle tissue** occurs in the walls of hollow organs such as the intestines, stomach, and urinary bladder, and around passages such as the respiratory tract and blood vessels. Smooth muscle has no striations, is not under voluntary control, has only one nucleus per cell, is tapered at both ends, and is called involuntary muscle.

**Cardiac muscle tissue** is only found in the heart, and cardiac contractions pump blood throughout the body and maintain blood pressure. Like skeletal muscle, cardiac muscle is striated, but unlike skeletal muscle, cardiac muscle cannot be consciously controlled and is called involuntary muscle. It has one nucleus per cell, is branched, and is distinguished by the presence of intercalated disks.

## **Skeletal Muscle Fiber Structure**

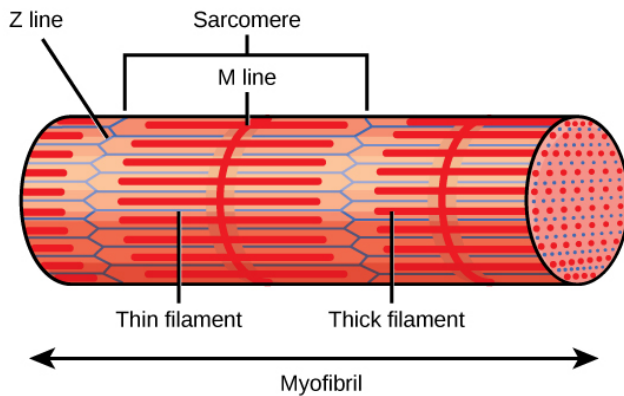
Each skeletal muscle fiber is a skeletal muscle cell. These cells are incredibly large, with diameters of up to 100  $\mu\text{m}$  and lengths of up to 30 cm. The plasma membrane of a skeletal muscle fiber is called the **sarcolemma**. The sarcolemma is the site of action potential conduction, which triggers muscle contraction. Within each muscle fiber are **myofibrils**—long cylindrical structures that lie parallel to the muscle fiber. Myofibrils run the entire length of the muscle fiber, and because they are only approximately 1.2  $\mu\text{m}$  in diameter, hundreds to thousands can be found inside one muscle fiber. They attach to the sarcolemma at their ends, so that as myofibrils shorten, the entire muscle cell contracts ([link](#)).



A skeletal muscle cell is surrounded by a plasma membrane called the sarcolemma with a cytoplasm called the sarcoplasm. A muscle fiber is composed of many fibrils, packaged into orderly units.

The striated appearance of skeletal muscle tissue is a result of repeating bands of the proteins actin and myosin that are present along the length of myofibrils. Dark A bands and light I bands repeat along myofibrils, and the alignment of myofibrils in the cell causes the entire cell to appear striated or banded.

Each I band has a dense line running vertically through the middle called a Z disc or Z line. The Z discs mark the border of units called **sarcomeres**, which are the functional units of skeletal muscle. One sarcomere is the space between two consecutive Z discs and contains one entire A band and two halves of an I band, one on either side of the A band. A myofibril is composed of many sarcomeres running along its length, and as the sarcomeres individually contract, the myofibrils and muscle cells shorten ([link](#)).



A sarcomere is the region from one Z line to the next Z line. Many sarcomeres are present in a myofibril, resulting in the striation pattern characteristic of skeletal muscle.

Myofibrils are composed of smaller structures called **myofilaments**. There are two main types of filaments: thick filaments and thin filaments; each has different compositions and locations. **Thick filaments** occur only in the A band of a myofibril. **Thin filaments** attach to a protein in the Z disc called alpha-actinin and occur across the entire length of the I band and partway into the A band. The region at which thick and thin filaments overlap has a dense appearance, as there is little space between the filaments. Thin filaments do not extend all the way into the A bands, leaving a central region of the A band that only contains thick filaments. This central region of the A band looks slightly lighter than the rest of the A band and is called the H zone. The middle of the H zone has a vertical line called the M line, at which accessory proteins hold together thick filaments. Both the Z disc and the M line hold myofilaments in place to maintain the structural arrangement and layering of the myofibril. Myofibrils are connected to each other by intermediate, or desmin, filaments that attach to the Z disc.

Thick and thin filaments are themselves composed of proteins. Thick filaments are composed of the protein myosin. The tail of a myosin molecule connects with other myosin molecules to form the central region of a thick filament near the M line, whereas the heads align on either side of the thick filament where the thin filaments overlap. The primary component of thin filaments is the actin protein. Two other components of the thin filament are tropomyosin and troponin. Actin has binding sites for myosin attachment. Strands of tropomyosin block the binding sites and prevent actin–myosin interactions when the muscles are at rest. Troponin consists of three globular subunits. One subunit binds to tropomyosin, one subunit binds to actin, and one subunit binds  $\text{Ca}^{2+}$  ions.

**Note:**

Link to Learning



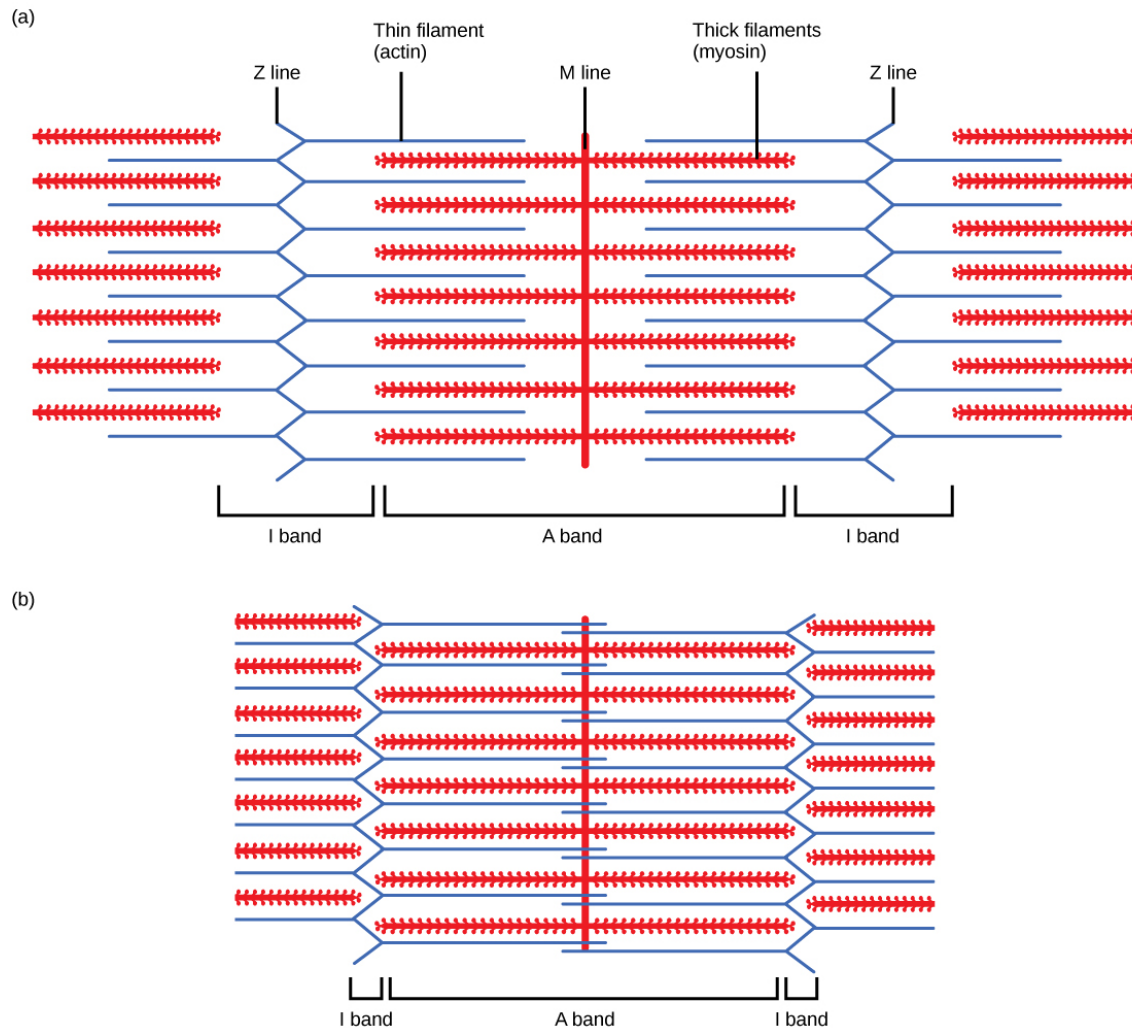
View this animation showing the organization of muscle fibers.

[https://www.openstaxcollege.org/l/skeletal\\_muscle](https://www.openstaxcollege.org/l/skeletal_muscle)

## Sliding Filament Model of Contraction

For a muscle cell to contract, the sarcomere must shorten. However, thick and thin filaments—the components of sarcomeres—do not shorten. Instead, they slide by one another, causing the sarcomere to shorten while the filaments remain the same length. The sliding filament theory of muscle contraction was developed to fit the differences observed in the named bands on the sarcomere at different degrees of muscle contraction and relaxation. The mechanism of contraction is the binding of myosin to actin, forming cross-bridges that generate filament movement ([link](#)).





When (a) a sarcomere (b) contracts, the Z lines move closer together and the I band gets smaller. The A band stays the same width and, at full contraction, the thin filaments overlap.

When a sarcomere shortens, some regions shorten whereas others stay the same length. A sarcomere is defined as the distance between two consecutive Z discs or Z lines; when a muscle contracts, the distance between the Z discs is reduced. The H zone—the central region of the A zone—contains only thick filaments and is shortened during contraction. The I band contains only thin filaments and also shortens. The A band does not shorten—it remains the same length—but A bands of different sarcomeres move closer together during contraction, eventually

disappearing. Thin filaments are pulled by the thick filaments toward the center of the sarcomere until the Z discs approach the thick filaments. The zone of overlap, in which thin filaments and thick filaments occupy the same area, increases as the thin filaments move inward.

## **ATP and Muscle Contraction**

The motion of muscle shortening occurs as myosin heads bind to actin and pull the actin inwards. This action requires energy, which is provided by ATP. Myosin binds to actin at a binding site on the globular actin protein. Myosin has another binding site for ATP at which enzymatic activity hydrolyzes ATP to ADP, releasing an inorganic phosphate molecule and energy.

ATP binding causes myosin to release actin, allowing actin and myosin to detach from each other. After this happens, the newly bound ATP is converted to ADP and inorganic phosphate,  $P_i$ . The enzyme at the binding site on myosin is called ATPase. The energy released during ATP hydrolysis changes the angle of the myosin head into a “cocked” position. The myosin head is then in a position for further movement, possessing potential energy, but ADP and  $P_i$  are still attached. If actin binding sites are covered and unavailable, the myosin will remain in the high energy configuration with ATP hydrolyzed, but still attached.

If the actin binding sites are uncovered, a cross-bridge will form; that is, the myosin head spans the distance between the actin and myosin molecules.  $P_i$  is then released, allowing myosin to expend the stored energy as a conformational change. The myosin head moves toward the M line, pulling the actin along with it. As the actin is pulled, the filaments move approximately 10 nm toward the M line. This movement is called the power stroke, as it is the step at which force is produced. As the actin is pulled toward the M line, the sarcomere shortens and the muscle contracts.

When the myosin head is “cocked,” it contains energy and is in a high-energy configuration. This energy is expended as the myosin head moves through the power stroke; at the end of the power stroke, the myosin head is in a low-energy position. After the power stroke, ADP is released; however,

the cross-bridge formed is still in place, and actin and myosin are bound together. ATP can then attach to myosin, which allows the cross-bridge cycle to start again and further muscle contraction can occur ([link](#)).

**Note:**

Link to Learning

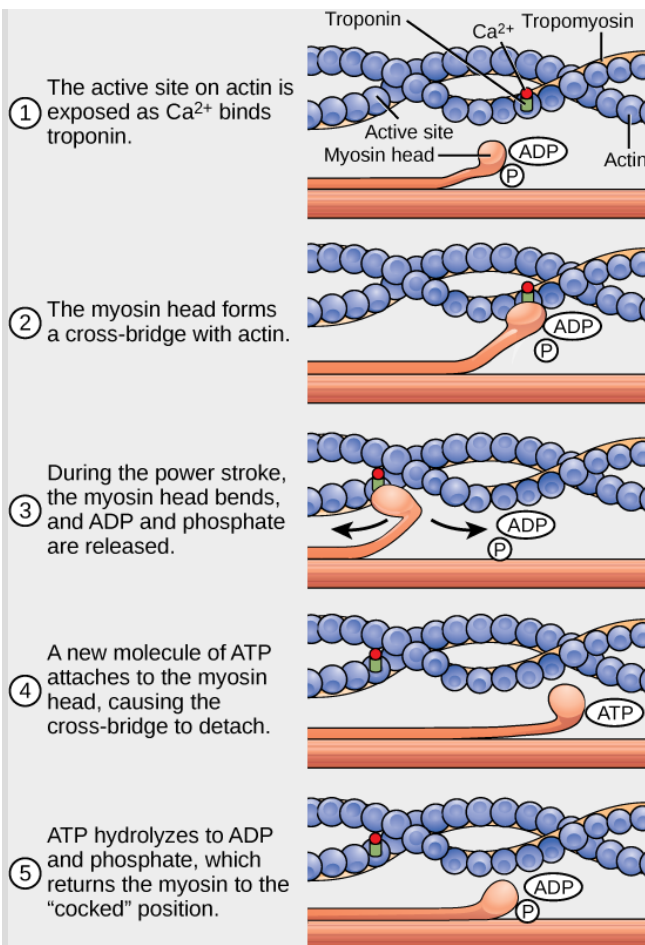


Watch this video explaining how a muscle contraction is signaled.

[https://www.openstaxcollege.org/l/contract\\_muscle](https://www.openstaxcollege.org/l/contract_muscle)

**Note:**

Art Connection



The cross-bridge muscle contraction cycle, which is triggered by  $\text{Ca}^{2+}$  binding to the actin active site, is shown. With each contraction cycle, actin moves relative to myosin.

Which of the following statements about muscle contraction is true?

- The power stroke occurs when ATP is hydrolyzed to ADP and phosphate.
- The power stroke occurs when ADP and phosphate dissociate from the myosin head.

- c. The power stroke occurs when ADP and phosphate dissociate from the actin active site.
- d. The power stroke occurs when  $\text{Ca}^{2+}$  binds the calcium head.

**Note:**

Link to Learning



View this [animation](#) of the cross-bridge muscle contraction.

## Regulatory Proteins

When a muscle is in a resting state, actin and myosin are separated. To keep actin from binding to the active site on myosin, regulatory proteins block the molecular binding sites. **Tropomyosin** blocks myosin binding sites on actin molecules, preventing cross-bridge formation and preventing contraction in a muscle without nervous input. **Troponin** binds to tropomyosin and helps to position it on the actin molecule; it also binds calcium ions.

To enable a muscle contraction, tropomyosin must change conformation, uncovering the myosin-binding site on an actin molecule and allowing cross-bridge formation. This can only happen in the presence of calcium, which is kept at extremely low concentrations in the sarcoplasm. If present, calcium ions bind to troponin, causing conformational changes in troponin that allow tropomyosin to move away from the myosin binding sites on actin. Once the tropomyosin is removed, a cross-bridge can form between actin and myosin, triggering contraction. Cross-bridge cycling continues

until  $\text{Ca}^{2+}$  ions and ATP are no longer available and tropomyosin again covers the binding sites on actin.

## Excitation–Contraction Coupling

Excitation–contraction coupling is the link (transduction) between the action potential generated in the sarcolemma and the start of a muscle contraction. The trigger for calcium release from the sarcoplasmic reticulum into the sarcoplasm is a neural signal. Each skeletal muscle fiber is controlled by a motor neuron, which conducts signals from the brain or spinal cord to the muscle. The area of the sarcolemma on the muscle fiber that interacts with the neuron is called the **motor end plate**. The end of the neuron's axon is called the synaptic terminal, and it does not actually contact the motor end plate. A small space called the synaptic cleft separates the synaptic terminal from the motor end plate. Electrical signals travel along the neuron's axon, which branches through the muscle and connects to individual muscle fibers at a neuromuscular junction.

The ability of cells to communicate electrically requires that the cells expend energy to create an electrical gradient across their cell membranes. This charge gradient is carried by ions, which are differentially distributed across the membrane. Each ion exerts an electrical influence and a concentration influence. Just as milk will eventually mix with coffee without the need to stir, ions also distribute themselves evenly, if they are permitted to do so. In this case, they are not permitted to return to an evenly mixed state.

The sodium–potassium ATPase uses cellular energy to move  $\text{K}^+$  ions inside the cell and  $\text{Na}^+$  ions outside. This alone accumulates a small electrical charge, but a big concentration gradient. There is lots of  $\text{K}^+$  in the cell and lots of  $\text{Na}^+$  outside the cell. Potassium is able to leave the cell through  $\text{K}^+$  channels that are open 90% of the time, and it does. However,  $\text{Na}^+$  channels are rarely open, so  $\text{Na}^+$  remains outside the cell. When  $\text{K}^+$  leaves the cell, obeying its concentration gradient, that effectively leaves a negative charge behind. So at rest, there is a large concentration gradient for  $\text{Na}^+$  to enter the cell, and there is an accumulation of negative charges left behind in the cell. This is the resting membrane potential. Potential in this context means a

separation of electrical charge that is capable of doing work. It is measured in volts, just like a battery. However, the transmembrane potential is considerably smaller (0.07 V); therefore, the small value is expressed as millivolts (mV) or 70 mV. Because the inside of a cell is negative compared with the outside, a minus sign signifies the excess of negative charges inside the cell,  $-70$  mV.

If an event changes the permeability of the membrane to  $\text{Na}^+$  ions, they will enter the cell. That will change the voltage. This is an electrical event, called an action potential, that can be used as a cellular signal.

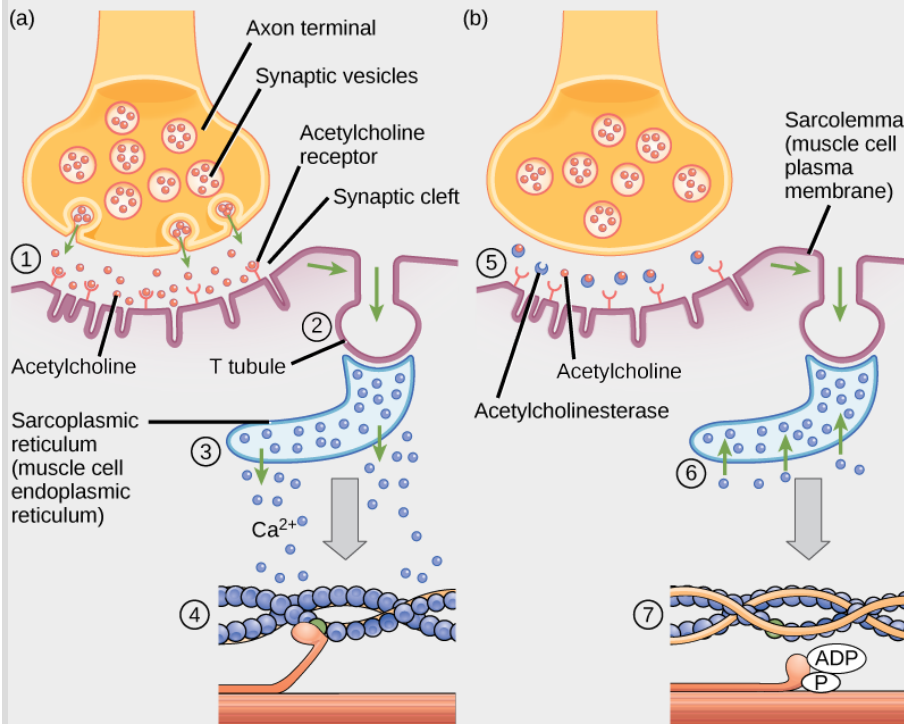
Communication occurs between nerves and muscles through neurotransmitters. Neuron action potentials cause the release of neurotransmitters from the synaptic terminal into the synaptic cleft, where they can then diffuse across the synaptic cleft and bind to a receptor molecule on the motor end plate. The motor end plate possesses junctional folds—folds in the sarcolemma that create a large surface area for the neurotransmitter to bind to receptors. The receptors are actually sodium channels that open to allow the passage of  $\text{Na}^+$  into the cell when they receive neurotransmitter signal.

Acetylcholine (ACh) is a neurotransmitter released by motor neurons that binds to receptors in the motor end plate. Neurotransmitter release occurs when an action potential travels down the motor neuron's axon, resulting in altered permeability of the synaptic terminal membrane and an influx of calcium. The  $\text{Ca}^{2+}$  ions allow synaptic vesicles to move to and bind with the presynaptic membrane (on the neuron), and release neurotransmitter from the vesicles into the synaptic cleft. Once released by the synaptic terminal, ACh diffuses across the synaptic cleft to the motor end plate, where it binds with ACh receptors. As a neurotransmitter binds, these ion channels open, and  $\text{Na}^+$  ions cross the membrane into the muscle cell. This reduces the voltage difference between the inside and outside of the cell, which is called depolarization. As ACh binds at the motor end plate, this depolarization is called an end-plate potential. The depolarization then spreads along the sarcolemma, creating an action potential as sodium channels adjacent to the initial depolarization site sense the change in voltage and open. The action potential moves across the entire cell, creating a wave of depolarization.

ACh is broken down by the enzyme **acetylcholinesterase** (AChE) into acetyl and choline. AChE resides in the synaptic cleft, breaking down ACh so that it does not remain bound to ACh receptors, which would cause unwanted extended muscle contraction ([\[link\]](#)).

### Note:

### Art Connection



1. Acetylcholine released from the axon terminal binds to receptors on the sarcolemma.
2. An action potential is generated and travels down the T tubule.
3.  $\text{Ca}^{2+}$  is released from the sarcoplasmic reticulum in response to the change in voltage.
4.  $\text{Ca}^{2+}$  binds troponin; Cross-bridges form between actin and myosin.

5. Acetylcholinesterase removes acetylcholine from the synaptic cleft.
6.  $\text{Ca}^{2+}$  is transported back into the sarcoplasmic reticulum.
7. Tropomyosin binds active sites on actin causing the cross-bridge to detach.

This diagram shows excitation-contraction coupling in a skeletal muscle contraction. The sarcoplasmic reticulum is a specialized endoplasmic reticulum found in muscle cells.



The deadly nerve gas Sarin irreversibly inhibits acetylcholinesterase. What effect would Sarin have on muscle contraction?

After depolarization, the membrane returns to its resting state. This is called repolarization, during which voltage-gated sodium channels close.

Potassium channels continue at 90% conductance. Because the plasma membrane sodium–potassium ATPase always transports ions, the resting state (negatively charged inside relative to the outside) is restored. The period immediately following the transmission of an impulse in a nerve or muscle, in which a neuron or muscle cell regains its ability to transmit another impulse, is called the refractory period. During the refractory period, the membrane cannot generate another action potential. . The refractory period allows the voltage-sensitive ion channels to return to their resting configurations. The sodium potassium ATPase continually moves  $\text{Na}^+$  back out of the cell and  $\text{K}^+$  back into the cell, and the  $\text{K}^+$  leaks out leaving negative charge behind. Very quickly, the membrane repolarizes, so that it can again be depolarized.

## **Control of Muscle Tension**

Neural control initiates the formation of actin–myosin cross-bridges, leading to the sarcomere shortening involved in muscle contraction. These contractions extend from the muscle fiber through connective tissue to pull on bones, causing skeletal movement. The pull exerted by a muscle is called tension, and the amount of force created by this tension can vary. This enables the same muscles to move very light objects and very heavy objects. In individual muscle fibers, the amount of tension produced depends on the cross-sectional area of the muscle fiber and the frequency of neural stimulation.

The number of cross-bridges formed between actin and myosin determine the amount of tension that a muscle fiber can produce. Cross-bridges can only form where thick and thin filaments overlap, allowing myosin to bind to actin. If more cross-bridges are formed, more myosin will pull on actin, and more tension will be produced.

The ideal length of a sarcomere during production of maximal tension occurs when thick and thin filaments overlap to the greatest degree. If a sarcomere at rest is stretched past an ideal resting length, thick and thin filaments do not overlap to the greatest degree, and fewer cross-bridges can form. This results in fewer myosin heads pulling on actin, and less tension is produced. As a sarcomere is shortened, the zone of overlap is reduced as the thin filaments reach the H zone, which is composed of myosin tails. Because it is myosin heads that form cross-bridges, actin will not bind to myosin in this zone, reducing the tension produced by this myofiber. If the sarcomere is shortened even more, thin filaments begin to overlap with each other—reducing cross-bridge formation even further, and producing even less tension. Conversely, if the sarcomere is stretched to the point at which thick and thin filaments do not overlap at all, no cross-bridges are formed and no tension is produced. This amount of stretching does not usually occur because accessory proteins, internal sensory nerves, and connective tissue oppose extreme stretching.

The primary variable determining force production is the number of myofibers within the muscle that receive an action potential from the neuron that controls that fiber. When using the biceps to pick up a pencil, the motor cortex of the brain only signals a few neurons of the biceps, and only a few myofibers respond. In vertebrates, each myofiber responds fully if stimulated. When picking up a piano, the motor cortex signals all of the neurons in the biceps and every myofiber participates. This is close to the maximum force the muscle can produce. As mentioned above, increasing the frequency of action potentials (the number of signals per second) can increase the force a bit more, because the tropomyosin is flooded with calcium.

## **Section Summary**

The body contains three types of muscle tissue: skeletal muscle, cardiac muscle, and smooth muscle. Skeletal muscle tissue is composed of sarcomeres, the functional units of muscle tissue. Muscle contraction occurs when sarcomeres shorten, as thick and thin filaments slide past each other, which is called the sliding filament model of muscle contraction. ATP provides the energy for cross-bridge formation and filament sliding.

Regulatory proteins, such as troponin and tropomyosin, control cross-bridge formation. Excitation–contraction coupling transduces the electrical signal of the neuron, via acetylcholine, to an electrical signal on the muscle membrane, which initiates force production. The number of muscle fibers contracting determines how much force the whole muscle produces.

## Art Connections

### Exercise:

#### Problem:

[\[link\]](#) Which of the following statements about muscle contraction is true?

- a. The power stroke occurs when ATP is hydrolyzed to ADP and phosphate.
- b. The power stroke occurs when ADP and phosphate dissociate from the myosin head.
- c. The power stroke occurs when ADP and phosphate dissociate from the actin active site.
- d. The power stroke occurs when  $\text{Ca}^{2+}$  binds the calcium head.

---

#### Solution:

[\[link\]](#) B

### Exercise:

#### Problem:

[\[link\]](#) The deadly nerve gas Sarin irreversibly inhibits acetylcholinesterase. What effect would Sarin have on muscle contraction?

---

#### Solution:

[\[link\]](#) In the presence of Sarin, acetylcholine is not removed from the synapse, resulting in continuous stimulation of the muscle plasma membrane. At first, muscle activity is intense and uncontrolled, but the ion gradients dissipate, so electrical signals in the T-tubules are no longer possible. The result is paralysis, leading to death by asphyxiation.

## Review Questions

### Exercise:

#### Problem:

In relaxed muscle, the myosin-binding site on actin is blocked by \_\_\_\_\_.

- a. titin
- b. troponin
- c. myoglobin
- d. tropomyosin

---

#### Solution:

D

### Exercise:

**Problem:** The cell membrane of a muscle fiber is called a \_\_\_\_\_.

- a. myofibril
- b. sarcolemma
- c. sarcoplasm
- d. myofilament

---

#### Solution:

B

**Exercise:**

**Problem:**

The muscle relaxes if no new nerve signal arrives. However the neurotransmitter from the previous stimulation is still present in the synapse. The activity of \_\_\_\_\_ helps to remove this neurotransmitter.

- a. myosin
- b. action potential
- c. tropomyosin
- d. acetylcholinesterase

---

**Solution:**

D

**Exercise:**

**Problem:**

The ability of a muscle to generate tension immediately after stimulation is dependent on:

- a. myosin interaction with the M line
- b. overlap of myosin and actin
- c. actin attachments to the Z line
- d. none of the above

---

**Solution:**

D

**Free Response**

**Exercise:****Problem:**

How would muscle contractions be affected if ATP was completely depleted in a muscle fiber?

---

**Solution:**

Because ATP is required for myosin to release from actin, muscles would remain rigidly contracted until more ATP was available for the myosin cross-bridge release. This is why dead vertebrates undergo rigor mortis.

**Exercise:****Problem:**

What factors contribute to the amount of tension produced in an individual muscle fiber?

---

**Solution:**

The cross-sectional area, the length of the muscle fiber at rest, and the frequency of neural stimulation.

**Exercise:****Problem:**

What effect will low blood calcium have on neurons? What effect will low blood calcium have on skeletal muscles?

---

**Solution:**

Neurons will not be able to release neurotransmitter without calcium. Skeletal muscles have calcium stored and don't need any from the outside.

**Glossary**

actin

globular contractile protein that interacts with myosin for muscle contraction

acetylcholinesterase

(AChE) enzyme that breaks down ACh into acetyl and choline

cardiac muscle

tissue muscle tissue found only in the heart; cardiac contractions pump blood throughout the body and maintain blood pressure

motor end plate

sarcolemma of the muscle fiber that interacts with the neuron

myofibril

long cylindrical structures that lie parallel to the muscle fiber

myofilament

small structures that make up myofibrils

myosin

contractile protein that interacts with actin for muscle contraction

sarcolemma

plasma membrane of a skeletal muscle fiber

sarcomere

functional unit of skeletal muscle

skeletal muscle tissue

forms skeletal muscles, which attach to bones and control locomotion and any movement that can be consciously controlled

smooth muscle

tissue occurs in the walls of hollow organs such as the intestines, stomach, and urinary bladder, and around passages such as the respiratory tract and blood vessels

thick filament

a group of myosin molecules

thin filament

two polymers of actin wound together along with tropomyosin and troponin

tropomyosin

acts to block myosin binding sites on actin molecules, preventing cross-bridge formation and preventing contraction until a muscle receives a neuron signal

troponin

binds to tropomyosin and helps to position it on the actin molecule, and also binds calcium ions



## Introduction

class="introduction"

An athlete's nervous system is hard at work during the planning and execution of a movement as precise as a high jump. Parts of the nervous system are involved in determining how hard to push off and when to turn, as well as controlling the muscles throughout the body that make this complicated movement possible without knocking the bar down—all

in just a few  
seconds.  
(credit:  
modification  
n of work  
by Shane T.  
McCoy,  
U.S. Navy)



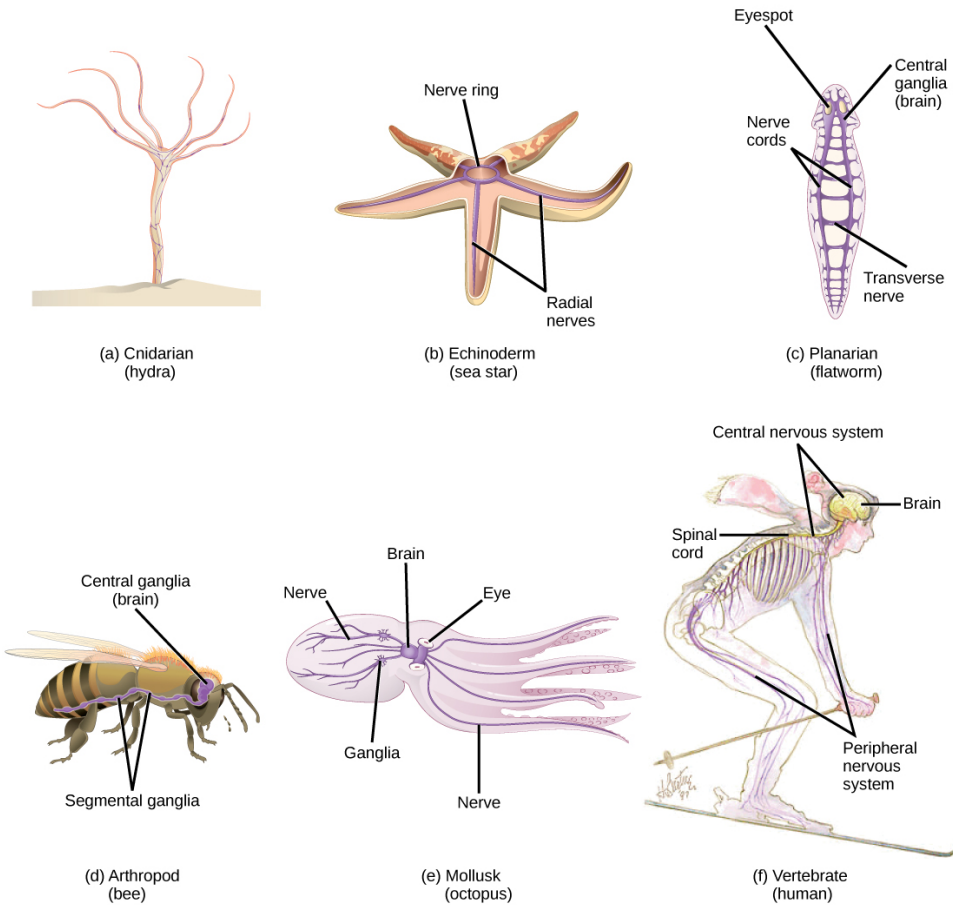
When you're reading this book, your nervous system is performing several functions simultaneously. The visual system is processing what is seen on the page; the motor system controls the turn of the pages (or click of the mouse); the prefrontal cortex maintains attention. Even fundamental functions, like breathing and regulation of body temperature, are controlled by the nervous system. A nervous system is an organism's control center: it processes sensory information from outside (and inside) the body and controls all behaviors—from eating to sleeping to finding a mate.

## Neurons and Glial Cells

By the end of this section, you will be able to:

- List and describe the functions of the structural components of a neuron
- List and describe the four main types of neurons
- Compare the functions of different types of glial cells

Nervous systems throughout the animal kingdom vary in structure and complexity, as illustrated by the variety of animals shown in [\[link\]](#). Some organisms, like sea sponges, lack a true nervous system. Others, like jellyfish, lack a true brain and instead have a system of separate but connected nerve cells (neurons) called a “nerve net.” Echinoderms such as sea stars have nerve cells that are bundled into fibers called nerves. Flatworms of the phylum Platyhelminthes have both a central nervous system (CNS), made up of a small “brain” and two nerve cords, and a peripheral nervous system (PNS) containing a system of nerves that extend throughout the body. The insect nervous system is more complex but also fairly decentralized. It contains a brain, ventral nerve cord, and ganglia (clusters of connected neurons). These ganglia can control movements and behaviors without input from the brain. Octopi may have the most complicated of invertebrate nervous systems—they have neurons that are organized in specialized lobes and eyes that are structurally similar to vertebrate species.



Nervous systems vary in structure and complexity. In (a) cnidarians, nerve cells form a decentralized nerve net. In (b) echinoderms, nerve cells are bundled into fibers called nerves. In animals exhibiting bilateral symmetry such as (c) planarians, neurons cluster into an anterior brain that processes information. In addition to a brain, (d) arthropods have clusters of nerve cell bodies, called peripheral ganglia, located along the ventral nerve cord. Mollusks such as squid and (e) octopi, which must hunt to survive, have complex brains containing millions of neurons. In (f) vertebrates, the brain and spinal cord comprise the central nervous system, while neurons extending into the rest of the body comprise the peripheral nervous system. (credit e: modification of work by Michael

Vecchione, Clyde F.E. Roper, and Michael J. Sweeney,  
NOAA; credit f: modification of work by NIH)

Compared to invertebrates, vertebrate nervous systems are more complex, centralized, and specialized. While there is great diversity among different vertebrate nervous systems, they all share a basic structure: a CNS that contains a brain and spinal cord and a PNS made up of peripheral sensory and motor nerves. One interesting difference between the nervous systems of invertebrates and vertebrates is that the nerve cords of many invertebrates are located ventrally whereas the vertebrate spinal cords are located dorsally. There is debate among evolutionary biologists as to whether these different nervous system plans evolved separately or whether the invertebrate body plan arrangement somehow “flipped” during the evolution of vertebrates.

**Note:**

Link to Learning



Watch this video of biologist Mark Kirschner discussing the “flipping” phenomenon of vertebrate evolution.

[https://www.openstaxcollege.org/l/vertebrate\\_evol](https://www.openstaxcollege.org/l/vertebrate_evol)

The nervous system is made up of **neurons**, specialized cells that can receive and transmit chemical or electrical signals, and **glia**, cells that provide support functions for the neurons by playing an information processing role that is complementary to neurons. A neuron can be compared to an electrical wire—it transmits a signal from one place to

another. Glia can be compared to the workers at the electric company who make sure wires go to the right places, maintain the wires, and take down wires that are broken. Although glia have been compared to workers, recent evidence suggests that also usurp some of the signaling functions of neurons.

There is great diversity in the types of neurons and glia that are present in different parts of the nervous system. There are four major types of neurons, and they share several important cellular components.

## Neurons

The nervous system of the common laboratory fly, *Drosophila melanogaster*, contains around 100,000 neurons, the same number as a lobster. This number compares to 75 million in the mouse and 300 million in the octopus. A human brain contains around 86 billion neurons. Despite these very different numbers, the nervous systems of these animals control many of the same behaviors—from basic reflexes to more complicated behaviors like finding food and courting mates. The ability of neurons to communicate with each other as well as with other types of cells underlies all of these behaviors.

Most neurons share the same cellular components. But neurons are also highly specialized—different types of neurons have different sizes and shapes that relate to their functional roles.

## Parts of a Neuron

Like other cells, each neuron has a cell body (or soma) that contains a nucleus, smooth and rough endoplasmic reticulum, Golgi apparatus, mitochondria, and other cellular components. Neurons also contain unique structures, illustrated in [\[link\]](#) for receiving and sending the electrical signals that make neuronal communication possible. **Dendrites** are tree-like structures that extend away from the cell body to receive messages from other neurons at specialized junctions called **synapses**. Although some neurons do not have any dendrites, some types of neurons have multiple

dendrites. Dendrites can have small protrusions called dendritic spines, which further increase surface area for possible synaptic connections.

Once a signal is received by the dendrite, it then travels passively to the cell body. The cell body contains a specialized structure, the **axon hillock** that integrates signals from multiple synapses and serves as a junction between the cell body and an **axon**. An axon is a tube-like structure that propagates the integrated signal to specialized endings called **axon terminals**. These terminals in turn synapse on other neurons, muscle, or target organs.

Chemicals released at axon terminals allow signals to be communicated to these other cells. Neurons usually have one or two axons, but some neurons, like amacrine cells in the retina, do not contain any axons. Some axons are covered with **myelin**, which acts as an insulator to minimize dissipation of the electrical signal as it travels down the axon, greatly increasing the speed on conduction. This insulation is important as the axon from a human motor neuron can be as long as a meter—from the base of the spine to the toes.

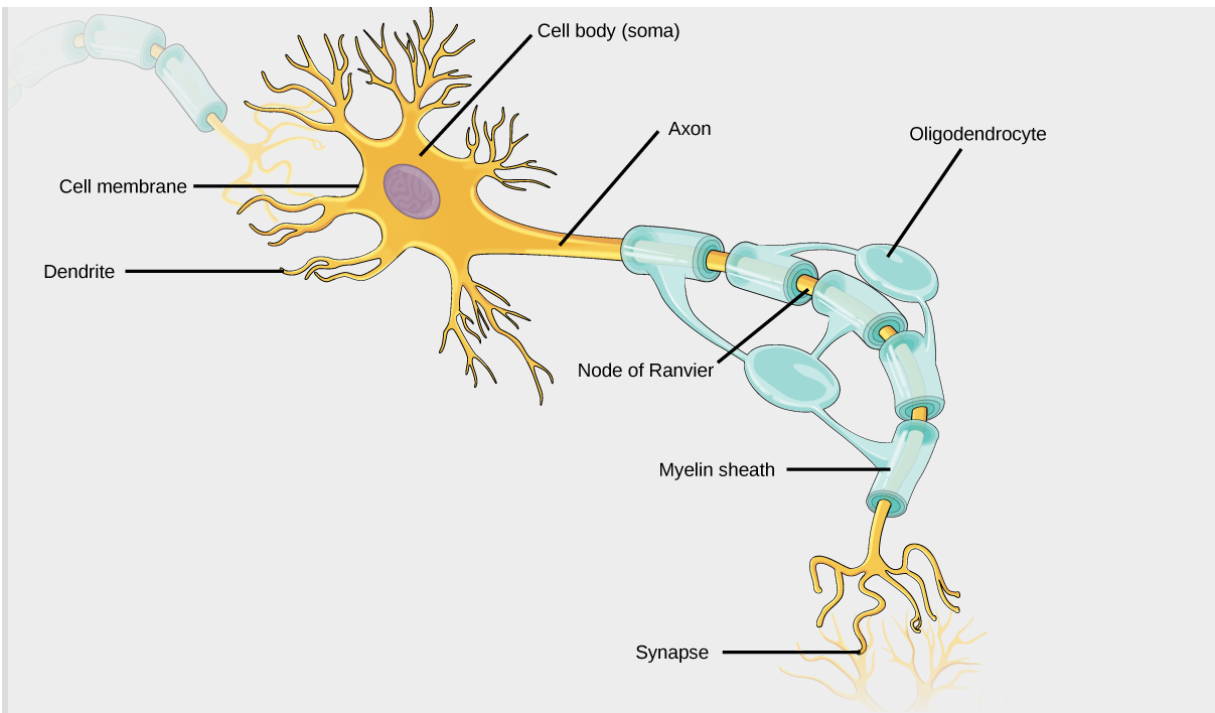
The myelin sheath is not actually part of the neuron. Myelin is produced by glial cells. Along the axon there are periodic gaps in the myelin sheath.

These gaps are called **nodes of Ranvier** and are sites where the signal is “recharged” as it travels along the axon.

It is important to note that a single neuron does not act alone—neuronal communication depends on the connections that neurons make with one another (as well as with other cells, like muscle cells). Dendrites from a single neuron may receive synaptic contact from many other neurons. For example, dendrites from a Purkinje cell in the cerebellum are thought to receive contact from as many as 200,000 other neurons.

### **Note:**

Art Connection



Neurons contain organelles common to many other cells, such as a nucleus and mitochondria. They also have more specialized structures, including dendrites and axons.

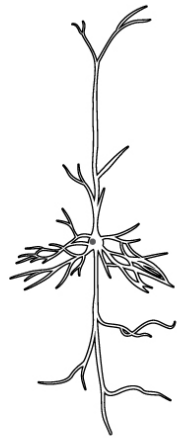
Which of the following statements is false?

- a. The soma is the cell body of a nerve cell.
- b. Myelin sheath provides an insulating layer to the dendrites.
- c. Axons carry the signal from the soma to the target.
- d. Dendrites carry the signal to the soma.

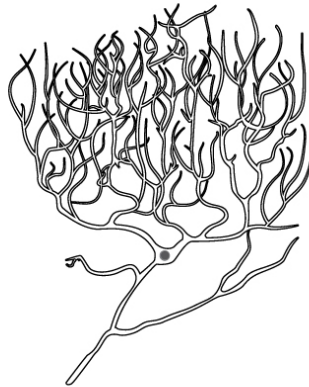
## Types of Neurons

There are different types of neurons, and the functional role of a given neuron is intimately dependent on its structure. There is an amazing diversity of neuron shapes and sizes found in different parts of the nervous system (and across species), as illustrated by the neurons shown in [\[link\]](#).

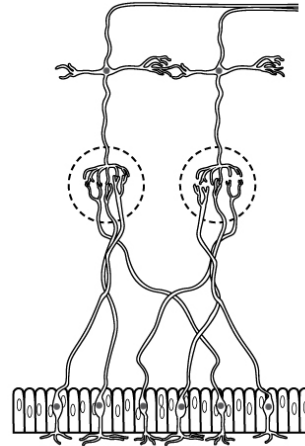




(a) Pyramidal cell of the cerebral cortex



(b) Purkinje cell of the cerebellar cortex

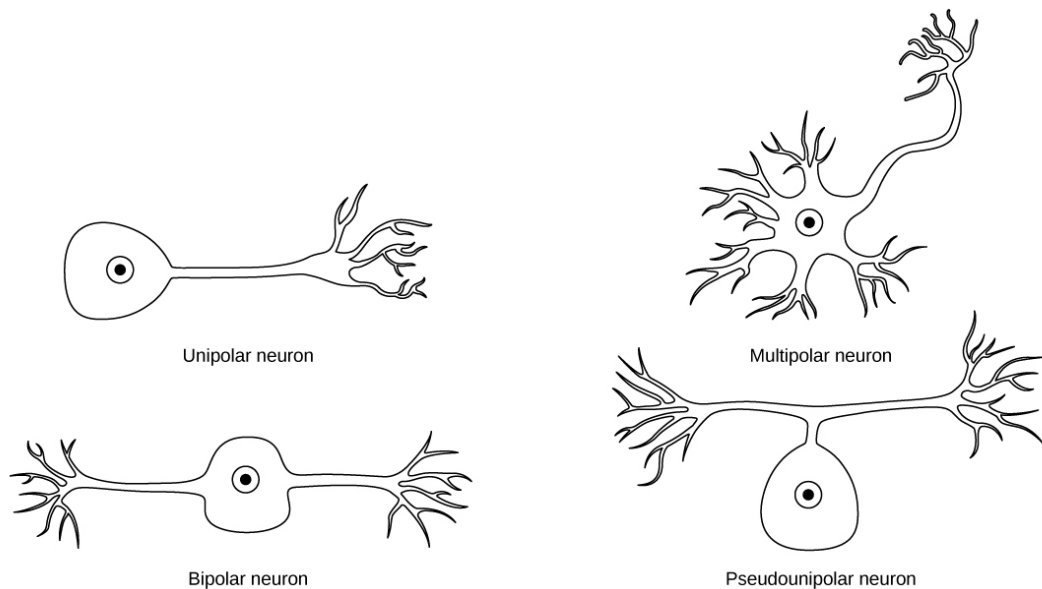


(c) Olfactory neurons

There is great diversity in the size and shape of neurons throughout the nervous system. Examples include (a) a pyramidal cell from the cerebral cortex, (b) a Purkinje cell from the cerebellar cortex, and (c) olfactory cells from the olfactory epithelium and olfactory bulb.

While there are many defined neuron cell subtypes, neurons are broadly divided into four basic types: unipolar, bipolar, multipolar, and pseudounipolar. [\[link\]](#) illustrates these four basic neuron types. Unipolar neurons have only one structure that extends away from the soma. These neurons are not found in vertebrates but are found in insects where they stimulate muscles or glands. A bipolar neuron has one axon and one dendrite extending from the soma. An example of a bipolar neuron is a retinal bipolar cell, which receives signals from photoreceptor cells that are sensitive to light and transmits these signals to ganglion cells that carry the signal to the brain. Multipolar neurons are the most common type of neuron. Each multipolar neuron contains one axon and multiple dendrites. Multipolar neurons can be found in the central nervous system (brain and spinal cord). An example of a multipolar neuron is a Purkinje cell in the cerebellum, which has many branching dendrites but only one axon. Pseudounipolar cells share characteristics with both unipolar and bipolar cells. A pseudounipolar cell has a single process that extends from the soma, like a unipolar cell, but this process later branches into two distinct

structures, like a bipolar cell. Most sensory neurons are pseudounipolar and have an axon that branches into two extensions: one connected to dendrites that receive sensory information and another that transmits this information to the spinal cord.



Neurons are broadly divided into four main types based on the number and placement of axons: (1) unipolar, (2) bipolar, (3) multipolar, and (4) pseudounipolar.

### **Note:**

#### **Everyday Connection**

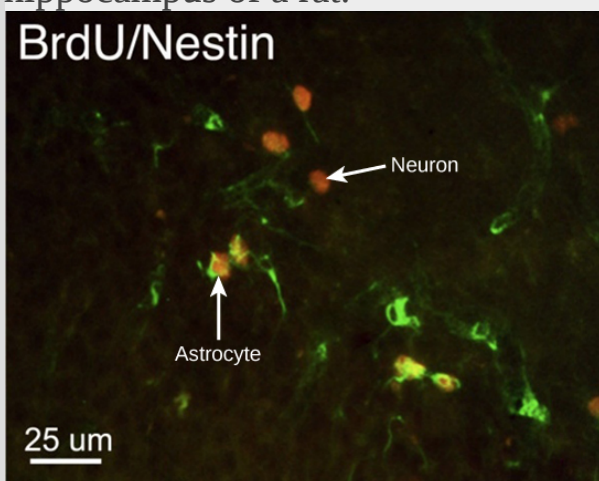
#### **Neurogenesis**

At one time, scientists believed that people were born with all the neurons they would ever have. Research performed during the last few decades indicates that neurogenesis, the birth of new neurons, continues into adulthood. Neurogenesis was first discovered in songbirds that produce new neurons while learning songs. For mammals, new neurons also play an important role in learning: about 1000 new neurons develop in the

hippocampus (a brain structure involved in learning and memory) each day. While most of the new neurons will die, researchers found that an increase in the number of surviving new neurons in the hippocampus correlated with how well rats learned a new task. Interestingly, both exercise and some antidepressant medications also promote neurogenesis in the hippocampus. Stress has the opposite effect. While neurogenesis is quite limited compared to regeneration in other tissues, research in this area may lead to new treatments for disorders such as Alzheimer's, stroke, and epilepsy.

How do scientists identify new neurons? A researcher can inject a compound called bromodeoxyuridine (BrdU) into the brain of an animal. While all cells will be exposed to BrdU, BrdU will only be incorporated into the DNA of newly generated cells that are in S phase. A technique called immunohistochemistry can be used to attach a fluorescent label to the incorporated BrdU, and a researcher can use fluorescent microscopy to visualize the presence of BrdU, and thus new neurons, in brain tissue.

[\[link\]](#) is a micrograph which shows fluorescently labeled neurons in the hippocampus of a rat.



This micrograph shows fluorescently labeled new neurons in a rat hippocampus. Cells that are actively dividing have bromodeoxyuridine (BrdU) incorporated into their DNA and are labeled in red. Cells that express glial fibrillary acidic

protein (GFAP) are labeled in green. Astrocytes, but not neurons, express GFAP. Thus, cells that are labeled both red and green are actively dividing astrocytes, whereas cells labeled red only are actively dividing neurons. (credit: modification of work by Dr. Maryam Faiz, et. al., University of Barcelona; scale-bar data from Matt Russell)

**Note:**

Link to Learning



[This site](#) contains more information about neurogenesis, including an interactive laboratory simulation and a video that explains how BrdU labels new cells.

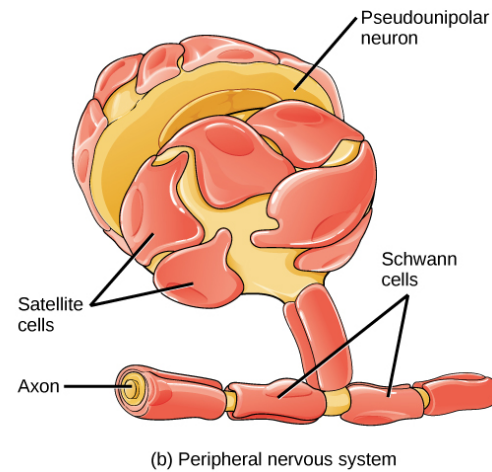
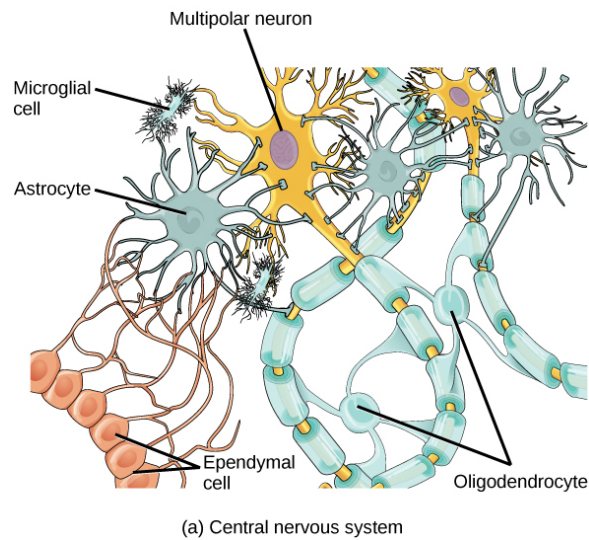
## Glia

While glia are often thought of as the supporting cast of the nervous system, the number of glial cells in the brain actually outnumbers the number of neurons by a factor of ten. Neurons would be unable to function without the

vital roles that are fulfilled by these glial cells. Glia guide developing neurons to their destinations, buffer ions and chemicals that would otherwise harm neurons, and provide myelin sheaths around axons. Scientists have recently discovered that they also play a role in responding to nerve activity and modulating communication between nerve cells. When glia do not function properly, the result can be disastrous—most brain tumors are caused by mutations in glia.

## Types of Glia

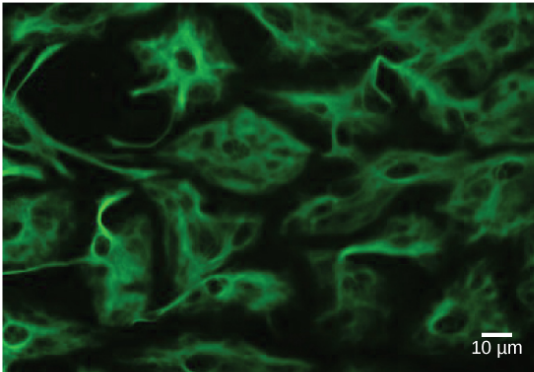
There are several different types of glia with different functions, two of which are shown in [\[link\]](#). **Astrocytes**, shown in [\[link\]a](#) make contact with both capillaries and neurons in the CNS. They provide nutrients and other substances to neurons, regulate the concentrations of ions and chemicals in the extracellular fluid, and provide structural support for synapses. Astrocytes also form the blood-brain barrier—a structure that blocks entrance of toxic substances into the brain. Astrocytes, in particular, have been shown through calcium imaging experiments to become active in response to nerve activity, transmit calcium waves between astrocytes, and modulate the activity of surrounding synapses. **Satellite glia** provide nutrients and structural support for neurons in the PNS. **Microglia** scavenge and degrade dead cells and protect the brain from invading microorganisms. **Oligodendrocytes**, shown in [\[link\]b](#) form myelin sheaths around axons in the CNS. One axon can be myelinated by several oligodendrocytes, and one oligodendrocyte can provide myelin for multiple neurons. This is distinctive from the PNS where a single **Schwann cell** provides myelin for only one axon as the entire Schwann cell surrounds the axon. **Radial glia** serve as scaffolds for developing neurons as they migrate to their end destinations. **Ependymal** cells line fluid-filled ventricles of the brain and the central canal of the spinal cord. They are involved in the production of cerebrospinal fluid, which serves as a cushion for the brain, moves the fluid between the spinal cord and the brain, and is a component for the choroid plexus.



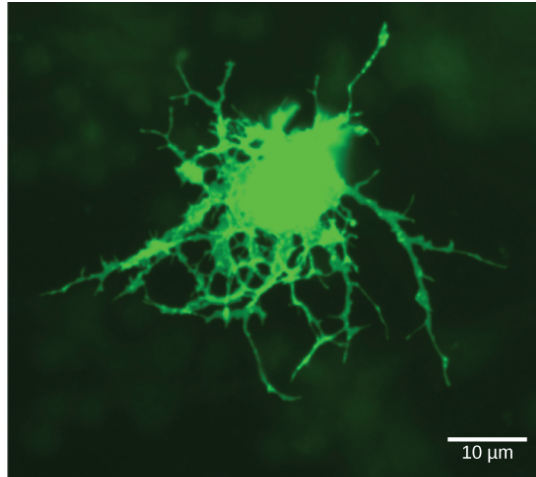
Glial cells support neurons and maintain their environment.

Glial cells of the (a) central nervous system include oligodendrocytes, astrocytes, ependymal cells, and microglial cells. Oligodendrocytes form the myelin sheath around axons.

Astrocytes provide nutrients to neurons, maintain their extracellular environment, and provide structural support. Microglia scavenge pathogens and dead cells. Ependymal cells produce cerebrospinal fluid that cushions the neurons. Glial cells of the (b) peripheral nervous system include Schwann cells, which form the myelin sheath, and satellite cells, which provide nutrients and structural support to neurons.



(a) Astrocyte



(b) Oligodendrocyte

(a) Astrocytes and (b) oligodendrocytes are glial cells of the central nervous system. (credit a: modification of work by Uniformed Services University; credit b: modification of work by Jurjen Broeke; scale-bar data from Matt Russell)

## Section Summary

The nervous system is made up of neurons and glia. Neurons are specialized cells that are capable of sending electrical as well as chemical signals. Most neurons contain dendrites, which receive these signals, and axons that send signals to other neurons or tissues. There are four main types of neurons: unipolar, bipolar, multipolar, and pseudounipolar neurons. Glia are non-neuronal cells in the nervous system that support neuronal development and signaling. There are several types of glia that serve different functions.

## Art Connections

### Exercise:

**Problem:** [\[link\]](#) Which of the following statements is false?

- a. The soma is the cell body of a nerve cell.
  - b. Myelin sheath provides an insulating layer to the dendrites.
  - c. Axons carry the signal from the soma to the target.
  - d. Dendrites carry the signal to the soma.
- 

**Solution:**

[\[link\]](#) B

## Review Questions

**Exercise:**

**Problem:**

Neurons contain \_\_\_\_\_, which can receive signals from other neurons.

- a. axons
  - b. mitochondria
  - c. dendrites
  - d. Golgi bodies
- 

**Solution:**

C

**Exercise:**

**Problem:**

A(n) \_\_\_\_\_ neuron has one axon and one dendrite extending directly from the cell body.

- a. unipolar
- b. bipolar
- c. multipolar



d. pseudounipolar

---

**Solution:**

B

**Exercise:**

**Problem:**

Glia that provide myelin for neurons in the brain are called \_\_\_\_\_.

- a. Schwann cells
- b. oligodendrocytes
- c. microglia
- d. astrocytes

---

**Solution:**

B

**Free Response**

**Exercise:**

**Problem:**

How are neurons similar to other cells? How are they unique?

---

**Solution:**

Neurons contain organelles common to all cells, such as a nucleus and mitochondria. They are unique because they contain dendrites, which can receive signals from other neurons, and axons that can send these signals to other cells.

**Exercise:**

**Problem:**

Multiple sclerosis causes demyelination of axons in the brain and spinal cord. Why is this problematic?

---

**Solution:**

Myelin provides insulation for signals traveling along axons. Without myelin, signal transmission can slow down and degrade over time. This would slow down neuronal communication across the nervous system and affect all downstream functions.

**Glossary**

astrocyte

glial cell in the central nervous system that provide nutrients, extracellular buffering, and structural support for neurons; also makes up the blood-brain barrier

axon

tube-like structure that propagates a signal from a neuron's cell body to axon terminals

axon hillock

electrically sensitive structure on the cell body of a neuron that integrates signals from multiple neuronal connections

axon terminal

structure on the end of an axon that can form a synapse with another neuron

dendrite

structure that extends away from the cell body to receive messages from other neurons

ependymal

cell that lines fluid-filled ventricles of the brain and the central canal of the spinal cord; involved in production of cerebrospinal fluid

glia

(also, glial cells) cells that provide support functions for neurons

microglia

glia that scavenge and degrade dead cells and protect the brain from invading microorganisms

myelin

fatty substance produced by glia that insulates axons

neuron

specialized cell that can receive and transmit electrical and chemical signals

nodes of Ranvier

gaps in the myelin sheath where the signal is recharged

oligodendrocyte

glial cell that myelinates central nervous system neuron axons

radial glia

glia that serve as scaffolds for developing neurons as they migrate to their final destinations

satellite glia

glial cell that provides nutrients and structural support for neurons in the peripheral nervous system

Schwann cell

glial cell that creates myelin sheath around a peripheral nervous system neuron axon

synapse

junction between two neurons where neuronal signals are communicated

## How Neurons Communicate

By the end of this section, you will be able to:

- Describe the basis of the resting membrane potential
- Explain the stages of an action potential and how action potentials are propagated
- Explain the similarities and differences between chemical and electrical synapses
- Describe long-term potentiation and long-term depression

All functions performed by the nervous system—from a simple motor reflex to more advanced functions like making a memory or a decision—require neurons to communicate with one another. While humans use words and body language to communicate, neurons use electrical and chemical signals. Just like a person in a committee, one neuron usually receives and synthesizes messages from multiple other neurons before “making the decision” to send the message on to other neurons.

## Nerve Impulse Transmission within a Neuron

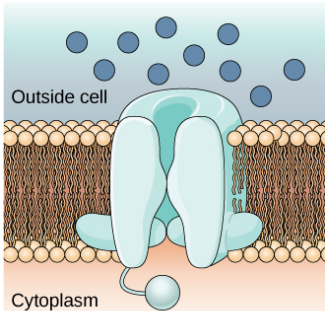
For the nervous system to function, neurons must be able to send and receive signals. These signals are possible because each neuron has a charged cellular membrane (a voltage difference between the inside and the outside), and the charge of this membrane can change in response to neurotransmitter molecules released from other neurons and environmental stimuli. To understand how neurons communicate, one must first understand the basis of the baseline or ‘resting’ membrane charge.

## Neuronal Charged Membranes

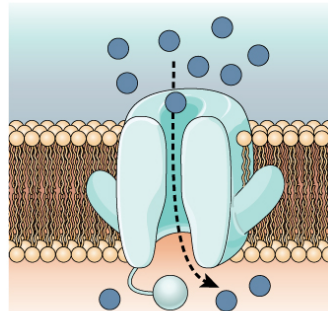
The lipid bilayer membrane that surrounds a neuron is impermeable to charged molecules or ions. To enter or exit the neuron, ions must pass through special proteins called ion channels that span the membrane. Ion channels have different configurations: open, closed, and inactive, as illustrated in [\[link\]](#). Some ion channels need to be activated in order to open and allow ions to pass into or out of the cell. These ion channels are

sensitive to the environment and can change their shape accordingly. Ion channels that change their structure in response to voltage changes are called voltage-gated ion channels. Voltage-gated ion channels regulate the relative concentrations of different ions inside and outside the cell. The difference in total charge between the inside and outside of the cell is called the **membrane potential**.

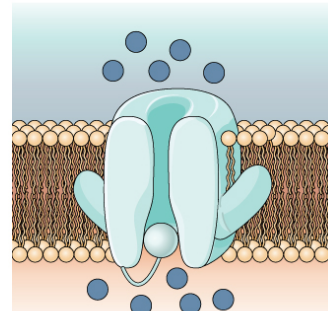
**Voltage-gated Na<sup>+</sup> Channels**



**Closed** At the resting potential, the channel is closed.



**Open** In response to a nerve impulse, the gate opens and Na<sup>+</sup> enters the cell.



**Inactivated** For a brief period following activation, the channel does not open in response to a new signal.

Voltage-gated ion channels open in response to changes in membrane voltage. After activation, they become inactivated for a brief period and will no longer open in response to a signal.

### Note:

Link to Learning



This video discusses the basis of the resting membrane potential.  
[https://www.openstaxcollege.org/l/resting\\_neuron](https://www.openstaxcollege.org/l/resting_neuron)

## Resting Membrane Potential

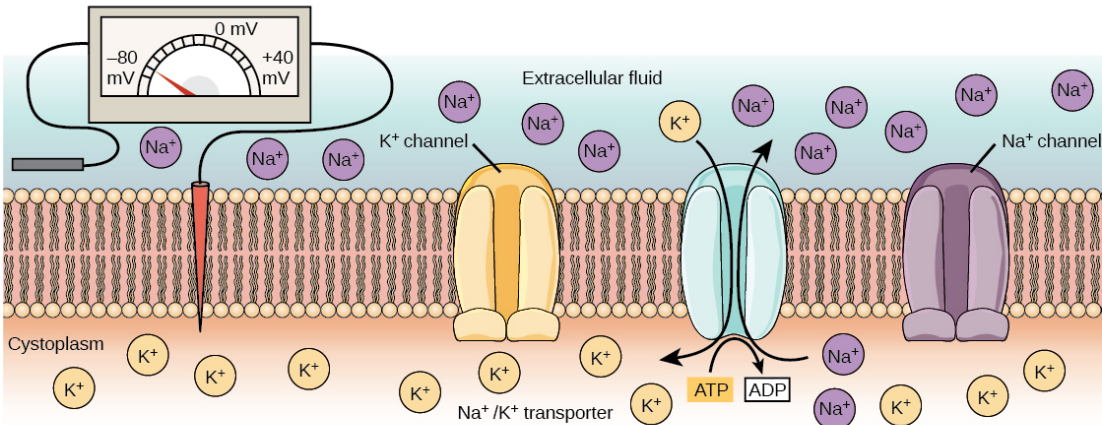
A neuron at rest is negatively charged: the inside of a cell is approximately 70 millivolts more negative than the outside ( $-70$  mV, note that this number varies by neuron type and by species). This voltage is called the resting membrane potential; it is caused by differences in the concentrations of ions inside and outside the cell. If the membrane were equally permeable to all ions, each type of ion would flow across the membrane and the system would reach equilibrium. Because ions cannot simply cross the membrane at will, there are different concentrations of several ions inside and outside the cell, as shown in [\[link\]](#). The difference in the number of positively charged potassium ions ( $K^+$ ) inside and outside the cell dominates the resting membrane potential ([\[link\]](#)). When the membrane is at rest,  $K^+$  ions accumulate inside the cell due to a net movement with the concentration gradient. The negative resting membrane potential is created and maintained by increasing the concentration of cations outside the cell (in the extracellular fluid) relative to inside the cell (in the cytoplasm). The negative charge within the cell is created by the cell membrane being more permeable to potassium ion movement than sodium ion movement. In neurons, potassium ions are maintained at high concentrations within the cell while sodium ions are maintained at high concentrations outside of the cell. The cell possesses potassium and sodium leakage channels that allow the two cations to diffuse down their concentration gradient. However, the neurons have far more potassium leakage channels than sodium leakage channels. Therefore, potassium diffuses out of the cell at a much faster rate than sodium leaks in. Because more cations are leaving the cell than are entering, this causes the interior of the cell to be negatively charged relative to the outside of the cell. The actions of the sodium potassium pump help to maintain the resting potential, once established. Recall that sodium potassium pumps bring two  $K^+$  ions into the cell while removing three  $Na^+$  ions per ATP consumed. As more cations are expelled from the cell than taken in, the inside of the cell remains negatively charged relative to the

extracellular fluid. It should be noted that chloride ions ( $\text{Cl}^-$ ) tend to accumulate outside of the cell because they are repelled by negatively-charged proteins within the cytoplasm.

Ion Concentration Inside and Outside Neurons			
Ion	Extracellular concentration (mM)	Intracellular concentration (mM)	Ratio outside/inside
$\text{Na}^+$	145	12	12
$\text{K}^+$	4	155	0.026
$\text{Cl}^-$	120	4	30
Organic anions ( $\text{A}^-$ )	—	100	

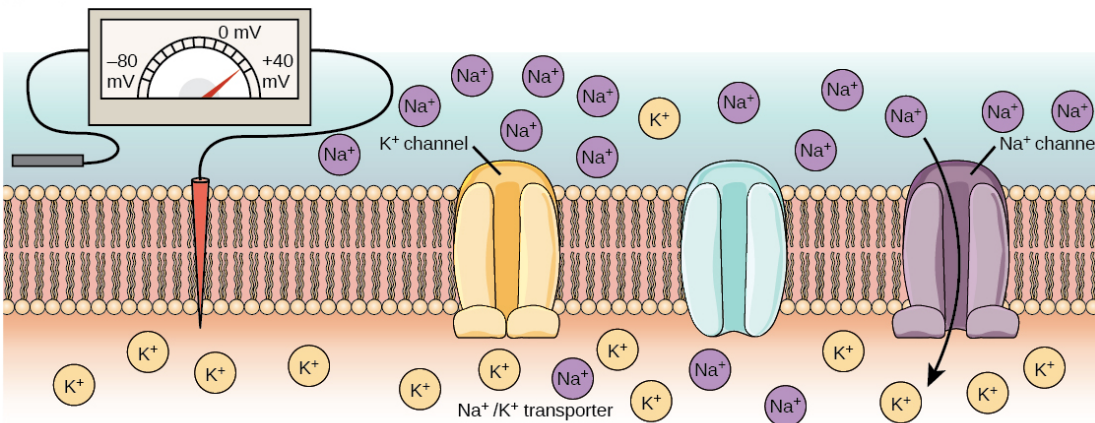
The resting membrane potential is a result of different concentrations inside and outside the cell.

(a) Resting potential



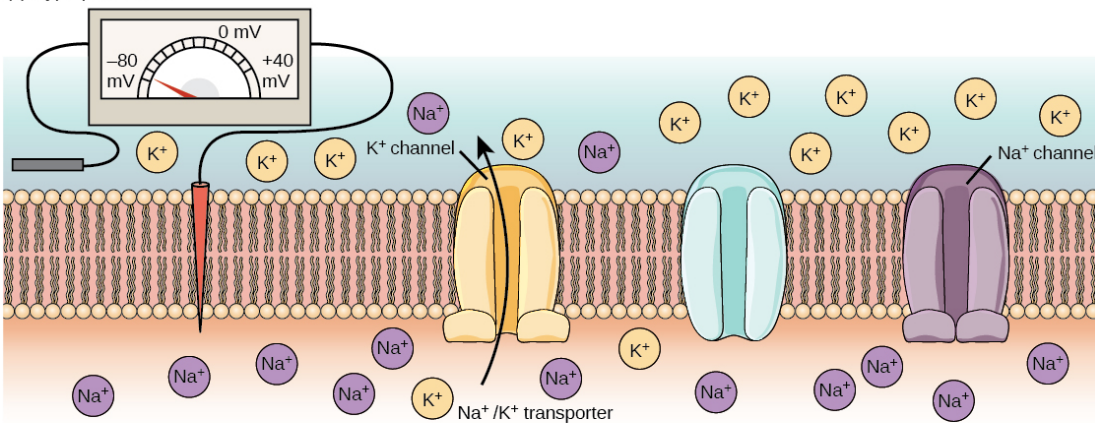
At the resting potential, all voltage-gated Na<sup>+</sup> channels and most voltage-gated K<sup>+</sup> channels are closed. The Na<sup>+</sup>/K<sup>+</sup> transporter pumps K<sup>+</sup> ions into the cell and Na<sup>+</sup> ions out.

(b) Depolarization



In response to a depolarization, some Na<sup>+</sup> channels open, allowing Na<sup>+</sup> ions to enter the cell. The membrane starts to depolarize (the charge across the membrane lessens). If the threshold of excitation is reached, all the Na<sup>+</sup> channels open.

(c) Hyperpolarization



At the peak action potential, Na<sup>+</sup> channels close while K<sup>+</sup> channels open. K<sup>+</sup> leaves the cell, and the membrane eventually becomes hyperpolarized.

The (a) resting membrane potential is a result of different concentrations of Na<sup>+</sup> and K<sup>+</sup> ions inside and outside the cell.



A nerve impulse causes  $\text{Na}^+$  to enter the cell, resulting in (b) depolarization. At the peak action potential,  $\text{K}^+$  channels open and the cell becomes (c) hyperpolarized.

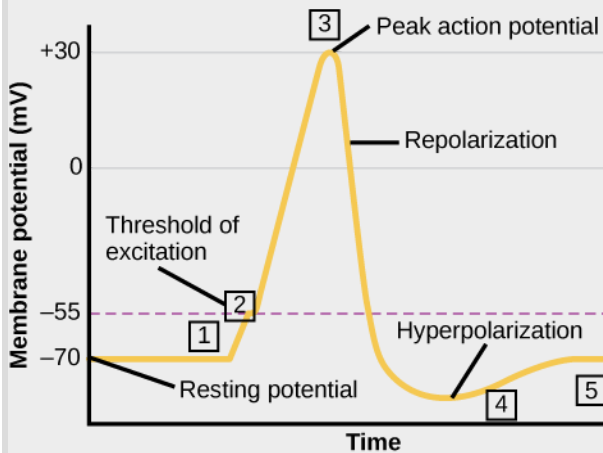
## Action Potential

A neuron can receive input from other neurons and, if this input is strong enough, send the signal to downstream neurons. Transmission of a signal between neurons is generally carried by a chemical called a neurotransmitter. Transmission of a signal within a neuron (from dendrite to axon terminal) is carried by a brief reversal of the resting membrane potential called an **action potential**. When neurotransmitter molecules bind to receptors located on a neuron's dendrites, ion channels open. At excitatory synapses, this opening allows positive ions to enter the neuron and results in **depolarization** of the membrane—a decrease in the difference in voltage between the inside and outside of the neuron. A stimulus from a sensory cell or another neuron depolarizes the target neuron to its threshold potential (-55 mV).  $\text{Na}^+$  channels in the axon hillock open, allowing positive ions to enter the cell ([\[link\]](#) and [\[link\]](#)). Once the sodium channels open, the neuron completely depolarizes to a membrane potential of about +40 mV. Action potentials are considered an "all-or nothing" event, in that, once the threshold potential is reached, the neuron always completely depolarizes. Once depolarization is complete, the cell must now "reset" its membrane voltage back to the resting potential. To accomplish this, the  $\text{Na}^+$  channels close and cannot be opened. This begins the neuron's **refractory period**, in which it cannot produce another action potential because its sodium channels will not open. At the same time, voltage-gated  $\text{K}^+$  channels open, allowing  $\text{K}^+$  to leave the cell. As  $\text{K}^+$  ions leave the cell, the membrane potential once again becomes negative. The diffusion of  $\text{K}^+$  out of the cell actually **hyperpolarizes** the cell, in that the membrane potential becomes more negative than the cell's normal resting potential. At this point, the sodium channels will return to their resting state, meaning they are ready to open again if the membrane potential again exceeds the

threshold potential. Eventually the extra  $K^+$  ions diffuse out of the cell through the potassium leakage channels, bringing the cell from its hyperpolarized state, back to its resting membrane potential.

**Note:**

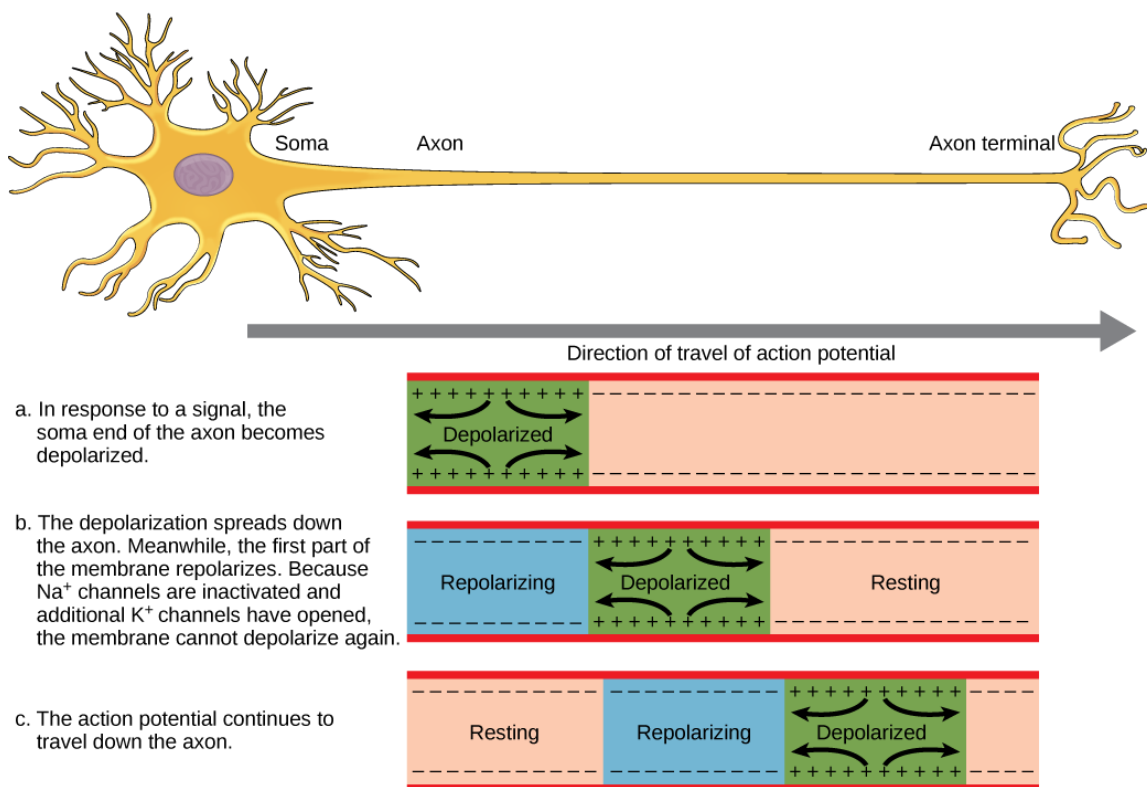
**Art Connection**



The formation of an action potential can be divided into five steps: (1) A stimulus from a sensory cell or another neuron causes the target cell to depolarize toward the threshold potential. (2) If the threshold of excitation is reached, all  $Na^+$  channels open and the membrane depolarizes. (3) At the peak action potential,  $K^+$  channels open and  $K^+$  begins to leave the cell. At the same time,  $Na^+$  channels close. (4) The membrane becomes hyperpolarized as  $K^+$  ions continue to leave the cell. The hyperpolarized membrane is in

a refractory period and cannot fire. (5) The  $K^+$  channels close and the  $Na^+/K^+$  transporter restores the resting potential.

Potassium channel blockers, such as amiodarone and procainamide, which are used to treat abnormal electrical activity in the heart, called cardiac dysrhythmia, impede the movement of  $K^+$  through voltage-gated  $K^+$  channels. Which part of the action potential would you expect potassium channels to affect?



The action potential is conducted down the axon as the axon membrane depolarizes, then repolarizes.

**Note:**

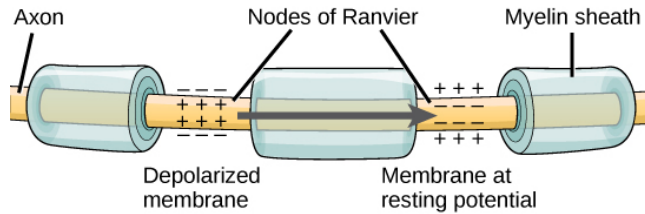
## Link to Learning



This [video](#) presents an overview of action potential.

## Myelin and the Propagation of the Action Potential

For an action potential to communicate information to another neuron, it must travel along the axon and reach the axon terminals where it can initiate neurotransmitter release. The speed of conduction of an action potential along an axon is influenced by both the diameter of the axon and the axon's resistance to current leak. Myelin acts as an insulator that prevents current from leaving the axon; this increases the speed of action potential conduction. In demyelinating diseases like multiple sclerosis, action potential conduction slows because current leaks from previously insulated axon areas. The nodes of Ranvier, illustrated in [\[link\]](#) are gaps in the myelin sheath along the axon. These unmyelinated spaces are about one micrometer long and contain voltage gated  $\text{Na}^+$  and  $\text{K}^+$  channels. Flow of ions through these channels, particularly the  $\text{Na}^+$  channels, regenerates the action potential over and over again along the axon. This 'jumping' of the action potential from one node to the next is called **saltatory conduction**. If nodes of Ranvier were not present along an axon, the action potential would propagate very slowly since  $\text{Na}^+$  and  $\text{K}^+$  channels would have to continuously regenerate action potentials at every point along the axon instead of at specific points. Nodes of Ranvier also save energy for the neuron since the channels only need to be present at the nodes and not along the entire axon.



Nodes of Ranvier are gaps in myelin coverage along axons. Nodes contain voltage-gated  $K^+$  and  $Na^+$  channels. Action potentials travel down the axon by jumping from one node to the next.

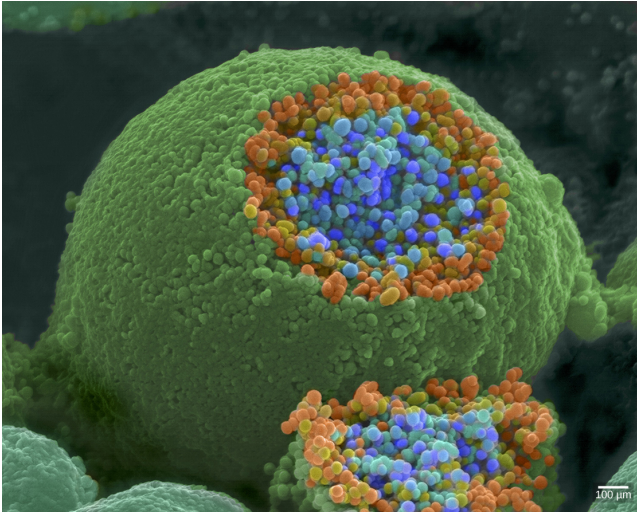
## Synaptic Transmission

The synapse or “gap” is the place where information is transmitted from one neuron to another. Synapses usually form between axon terminals and dendritic spines, but this is not universally true. There are also axon-to-axon, dendrite-to-dendrite, and axon-to-cell body synapses. The neuron transmitting the signal is called the presynaptic neuron, and the neuron receiving the signal is called the postsynaptic neuron. Note that these designations are relative to a particular synapse—most neurons are both presynaptic and postsynaptic. There are two types of synapses: chemical and electrical.

## Chemical Synapse

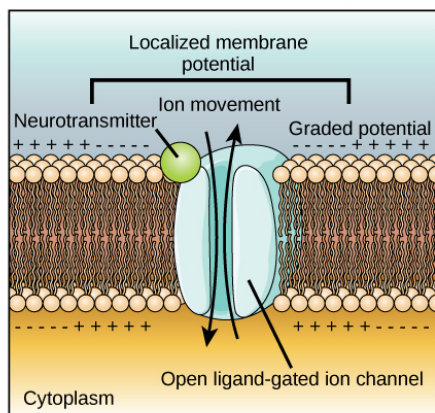
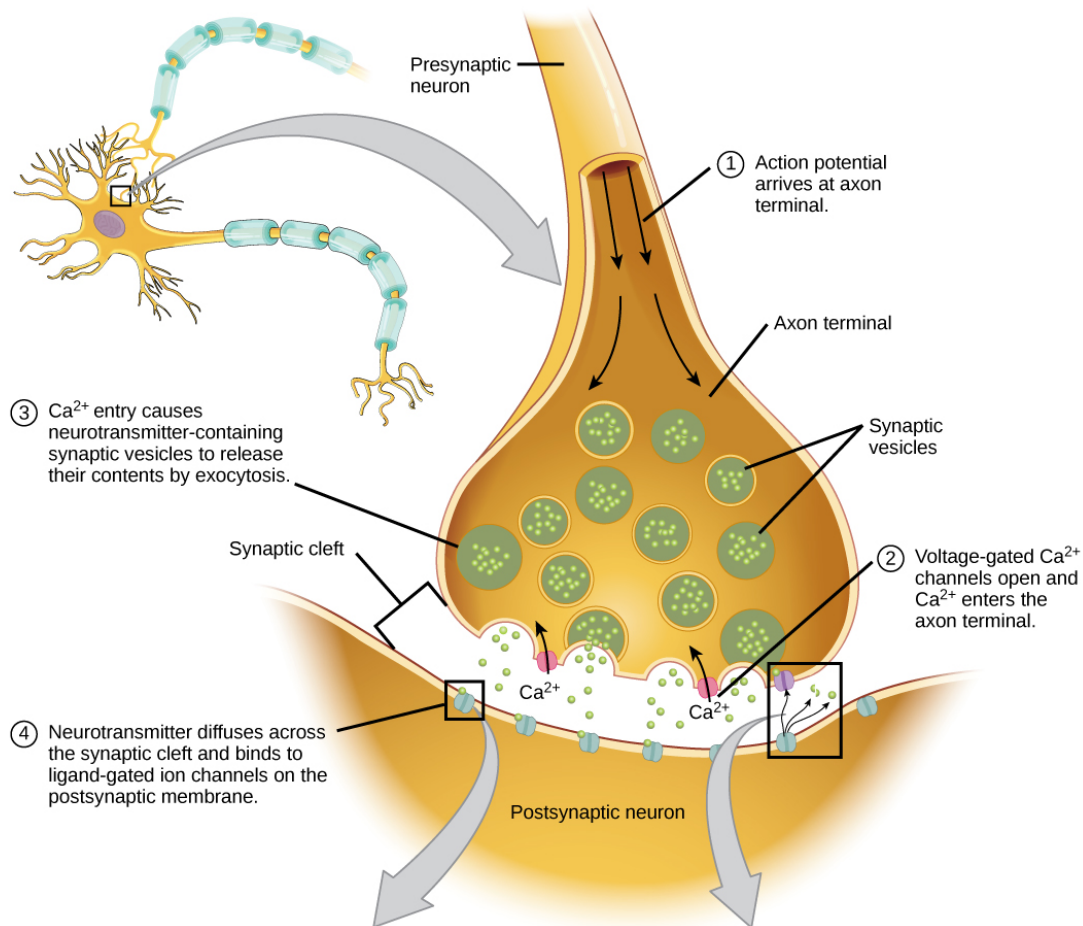
When an action potential reaches the axon terminal it depolarizes the membrane and opens voltage-gated  $Na^+$  channels.  $Na^+$  ions enter the cell, further depolarizing the presynaptic membrane. This depolarization causes voltage-gated  $Ca^{2+}$  channels to open. Calcium ions entering the cell initiate a signaling cascade that causes small membrane-bound vesicles, called

**synaptic vesicles**, containing neurotransmitter molecules to fuse with the presynaptic membrane. Synaptic vesicles are shown in [\[link\]](#), which is an image from a scanning electron microscope.

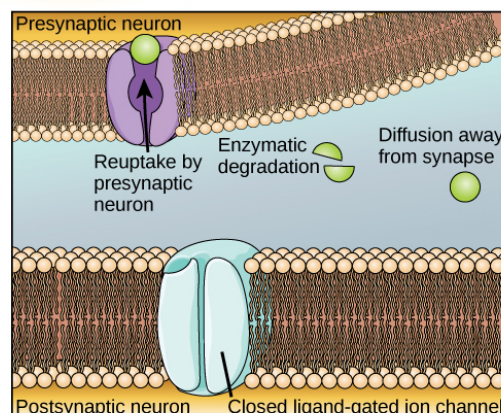


This pseudocolored image taken with a scanning electron microscope shows an axon terminal that was broken open to reveal synaptic vesicles (blue and orange) inside the neuron. (credit: modification of work by Tina Carvalho, NIH-NIGMS; scale-bar data from Matt Russell)

Fusion of a vesicle with the presynaptic membrane causes neurotransmitter to be released into the **synaptic cleft**, the extracellular space between the presynaptic and postsynaptic membranes, as illustrated in [\[link\]](#). The neurotransmitter diffuses across the synaptic cleft and binds to receptor proteins on the postsynaptic membrane.



- ⑤ Binding of neurotransmitter opens ligand-gated ion channels, resulting in graded potentials.



- ⑥ Reuptake by the presynaptic neuron, enzymatic degradation, and diffusion reduce neurotransmitter levels, terminating the signal.

Communication at chemical synapses requires release of neurotransmitters. When the presynaptic membrane is depolarized, voltage-gated  $\text{Ca}^{2+}$  channels open and allow  $\text{Ca}^{2+}$  to enter the cell. The calcium entry causes synaptic vesicles to fuse



with the membrane and release neurotransmitter molecules into the synaptic cleft. The neurotransmitter diffuses across the synaptic cleft and binds to ligand-gated ion channels in the postsynaptic membrane, resulting in a localized depolarization or hyperpolarization of the postsynaptic neuron.

The binding of a specific neurotransmitter causes particular ion channels, in this case ligand-gated channels, on the postsynaptic membrane to open. Neurotransmitters can either have excitatory or inhibitory effects on the postsynaptic membrane, as detailed in [\[link\]](#). For example, when acetylcholine is released at the synapse between a nerve and muscle (called the neuromuscular junction) by a presynaptic neuron, it causes postsynaptic  $\text{Na}^+$  channels to open.  $\text{Na}^+$  enters the postsynaptic cell and causes the postsynaptic membrane to depolarize. This depolarization is called an **excitatory postsynaptic potential (EPSP)** and makes the postsynaptic neuron more likely to fire an action potential. Release of neurotransmitter at inhibitory synapses causes **inhibitory postsynaptic potentials (IPSPs)**, a hyperpolarization of the postsynaptic membrane. For example, when the neurotransmitter GABA (gamma-aminobutyric acid) is released from a presynaptic neuron, it binds to and opens  $\text{Cl}^-$  channels.  $\text{Cl}^-$  ions enter the cell and hyperpolarizes the membrane, making the neuron less likely to fire an action potential.

Once neurotransmission has occurred, the neurotransmitter must be removed from the synaptic cleft so the postsynaptic membrane can “reset” and be ready to receive another signal. This can be accomplished in three ways: the neurotransmitter can diffuse away from the synaptic cleft, it can be degraded by enzymes in the synaptic cleft, or it can be recycled (sometimes called reuptake) by the presynaptic neuron. Several drugs act at this step of neurotransmission. For example, some drugs that are given to Alzheimer’s patients work by inhibiting acetylcholinesterase, the enzyme that degrades acetylcholine. This inhibition of the enzyme essentially increases neurotransmission at synapses that release acetylcholine. Once released, the acetylcholine stays in the cleft and can continually bind and unbind to postsynaptic receptors.



<b>Neurotransmitter Function and Location</b>		
<b>Neurotransmitter</b>	<b>Example</b>	<b>Location</b>
Acetylcholine	—	CNS and/or PNS
Biogenic amine	Dopamine, serotonin, norepinephrine	CNS and/or PNS
Amino acid	Glycine, glutamate, aspartate, gamma aminobutyric acid	CNS
Neuropeptide	Substance P, endorphins	CNS and/or PNS

## Electrical Synapse

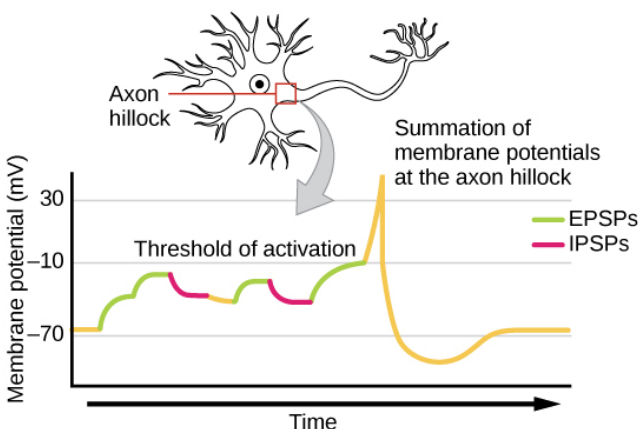
While electrical synapses are fewer in number than chemical synapses, they are found in all nervous systems and play important and unique roles. The mode of neurotransmission in electrical synapses is quite different from that in chemical synapses. In an electrical synapse, the presynaptic and postsynaptic membranes are very close together and are actually physically connected by channel proteins forming gap junctions. Gap junctions allow current to pass directly from one cell to the next. In addition to the ions that carry this current, other molecules, such as ATP, can diffuse through the large gap junction pores.

There are key differences between chemical and electrical synapses. Because chemical synapses depend on the release of neurotransmitter molecules from synaptic vesicles to pass on their signal, there is an approximately one millisecond delay between when the axon potential

reaches the presynaptic terminal and when the neurotransmitter leads to opening of postsynaptic ion channels. Additionally, this signaling is unidirectional. Signaling in electrical synapses, in contrast, is virtually instantaneous (which is important for synapses involved in key reflexes), and some electrical synapses are bidirectional. Electrical synapses are also more reliable as they are less likely to be blocked, and they are important for synchronizing the electrical activity of a group of neurons. For example, electrical synapses in the thalamus are thought to regulate slow-wave sleep, and disruption of these synapses can cause seizures.

## Signal Summation

Sometimes a single EPSP is strong enough to induce an action potential in the postsynaptic neuron, but often multiple presynaptic inputs must create EPSPs around the same time for the postsynaptic neuron to be sufficiently depolarized to fire an action potential. This process is called **summation** and occurs at the axon hillock, as illustrated in [\[link\]](#). Additionally, one neuron often has inputs from many presynaptic neurons—some excitatory and some inhibitory—so IPSPs can cancel out EPSPs and vice versa. It is the net change in postsynaptic membrane voltage that determines whether the postsynaptic cell has reached its threshold of excitation needed to fire an action potential. Together, synaptic summation and the threshold for excitation act as a filter so that random “noise” in the system is not transmitted as important information.



A single neuron can receive both excitatory and inhibitory inputs from multiple neurons, resulting in local membrane depolarization (EPSP input) and hyperpolarization (IPSP input). All these inputs are added together at the axon hillock. If the EPSPs are strong enough to overcome the IPSPs and reach the threshold of excitation, the neuron will fire.

**Note:**

**Everyday Connection**

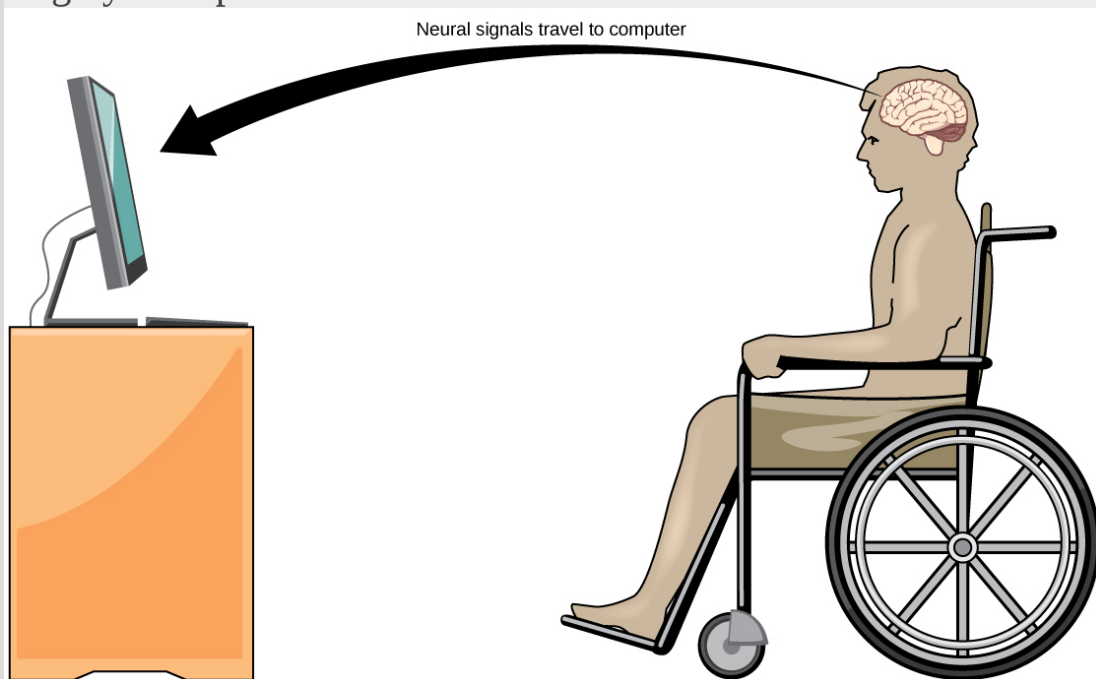
**Brain-computer interface**

Amyotrophic lateral sclerosis (ALS, also called Lou Gehrig's Disease) is a neurological disease characterized by the degeneration of the motor neurons that control voluntary movements. The disease begins with muscle weakening and lack of coordination and eventually destroys the neurons that control speech, breathing, and swallowing; in the end, the disease can lead to paralysis. At that point, patients require assistance from machines to be able to breathe and to communicate. Several special technologies have been developed to allow "locked-in" patients to communicate with the rest of the world. One technology, for example, allows patients to type out sentences by twitching their cheek. These sentences can then be read aloud by a computer.

A relatively new line of research for helping paralyzed patients, including those with ALS, to communicate and retain a degree of self-sufficiency is called brain-computer interface (BCI) technology and is illustrated in [\[link\]](#). This technology sounds like something out of science fiction: it allows paralyzed patients to control a computer using only their thoughts. There are several forms of BCI. Some forms use EEG recordings from electrodes taped onto the skull. These recordings contain information from

large populations of neurons that can be decoded by a computer. Other forms of BCI require the implantation of an array of electrodes smaller than a postage stamp in the arm and hand area of the motor cortex. This form of BCI, while more invasive, is very powerful as each electrode can record actual action potentials from one or more neurons. These signals are then sent to a computer, which has been trained to decode the signal and feed it to a tool—such as a cursor on a computer screen. This means that a patient with ALS can use e-mail, read the Internet, and communicate with others by thinking of moving his or her hand or arm (even though the paralyzed patient cannot make that bodily movement). Recent advances have allowed a paralyzed locked-in patient who suffered a stroke 15 years ago to control a robotic arm and even to feed herself coffee using BCI technology.

Despite the amazing advancements in BCI technology, it also has limitations. The technology can require many hours of training and long periods of intense concentration for the patient; it can also require brain surgery to implant the devices.



With brain-computer interface technology, neural signals from a paralyzed patient are collected, decoded, and then fed to a tool, such as a computer, a wheelchair, or a robotic arm.

### Note:

#### Link to Learning



Watch [this video](https://www.openstaxcollege.org/l/paralyzation) in which a paralyzed woman use a brain-controlled robotic arm to bring a drink to her mouth, among other images of brain-computer interface technology in action.

<https://www.openstaxcollege.org/l/paralyzation>

## Synaptic Plasticity

Synapses are not static structures. They can be weakened or strengthened. They can be broken, and new synapses can be made. Synaptic plasticity allows for these changes, which are all needed for a functioning nervous system. In fact, synaptic plasticity is the basis of learning and memory. Two processes in particular, long-term potentiation (LTP) and long-term depression (LTD) are important forms of synaptic plasticity that occur in synapses in the hippocampus, a brain region that is involved in storing memories.

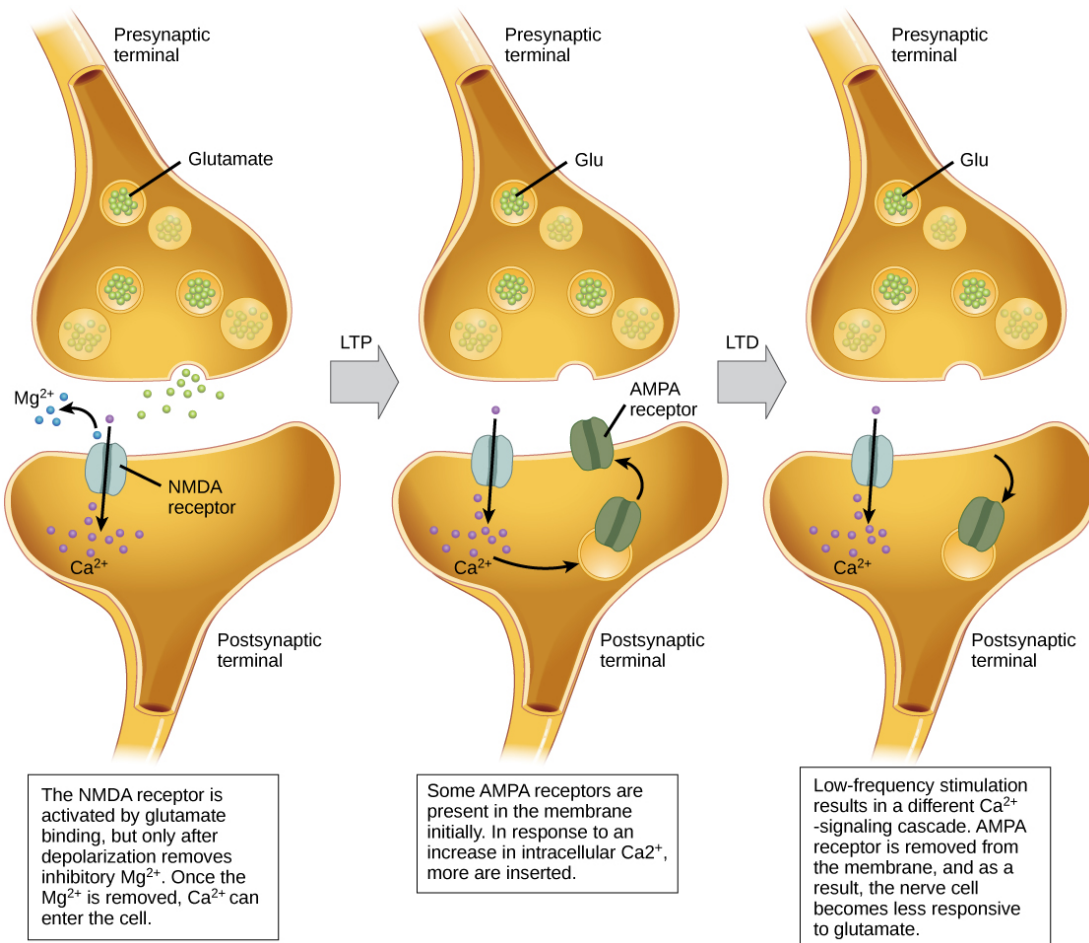
### Long-term Potentiation (LTP)

**Long-term potentiation (LTP)** is a persistent strengthening of a synaptic connection. LTP is based on the Hebbian principle: cells that fire together wire together. There are various mechanisms, none fully understood, behind the synaptic strengthening seen with LTP. One known mechanism involves

a type of postsynaptic glutamate receptor, called NMDA (N-Methyl-D-aspartate) receptors, shown in [\[link\]](#). These receptors are normally blocked by magnesium ions; however, when the postsynaptic neuron is depolarized by multiple presynaptic inputs in quick succession (either from one neuron or multiple neurons), the magnesium ions are forced out allowing Ca ions to pass into the postsynaptic cell. Next,  $\text{Ca}^{2+}$  ions entering the cell initiate a signaling cascade that causes a different type of glutamate receptor, called AMPA ( $\alpha$ -amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid) receptors, to be inserted into the postsynaptic membrane, since activated AMPA receptors allow positive ions to enter the cell. So, the next time glutamate is released from the presynaptic membrane, it will have a larger excitatory effect (EPSP) on the postsynaptic cell because the binding of glutamate to these AMPA receptors will allow more positive ions into the cell. The insertion of additional AMPA receptors strengthens the synapse and means that the postsynaptic neuron is more likely to fire in response to presynaptic neurotransmitter release. Some drugs of abuse co-opt the LTP pathway, and this synaptic strengthening can lead to addiction.

### **Long-term Depression (LTD)**

**Long-term depression (LTD)** is essentially the reverse of LTP: it is a long-term weakening of a synaptic connection. One mechanism known to cause LTD also involves AMPA receptors. In this situation, calcium that enters through NMDA receptors initiates a different signaling cascade, which results in the removal of AMPA receptors from the postsynaptic membrane, as illustrated in [\[link\]](#). The decrease in AMPA receptors in the membrane makes the postsynaptic neuron less responsive to glutamate released from the presynaptic neuron. While it may seem counterintuitive, LTD may be just as important for learning and memory as LTP. The weakening and pruning of unused synapses allows for unimportant connections to be lost and makes the synapses that have undergone LTP that much stronger by comparison.



Calcium entry through postsynaptic NMDA receptors can initiate two different forms of synaptic plasticity: long-term potentiation (LTP) and long-term depression (LTD). LTP arises when a single synapse is repeatedly stimulated. This stimulation causes a calcium- and CaMKII-dependent cellular cascade, which results in the insertion of more AMPA receptors into the postsynaptic membrane. The next time glutamate is released from the presynaptic cell, it will bind to both NMDA and the newly inserted AMPA receptors, thus depolarizing the membrane more efficiently. LTD occurs when few glutamate molecules bind to NMDA receptors at a synapse (due to a low firing rate of the presynaptic neuron). The calcium that does flow through NMDA receptors initiates a different calcineurin and protein phosphatase 1-dependent cascade, which results in the endocytosis of AMPA receptors.

This makes the postsynaptic neuron less responsive to glutamate released from the presynaptic neuron.

## Section Summary

Neurons have charged membranes because there are different concentrations of ions inside and outside of the cell. Voltage-gated ion channels control the movement of ions into and out of a neuron. When a neuronal membrane is depolarized to at least the threshold of excitation, an action potential is fired. The action potential is then propagated along a myelinated axon to the axon terminals. In a chemical synapse, the action potential causes release of neurotransmitter molecules into the synaptic cleft. Through binding to postsynaptic receptors, the neurotransmitter can cause excitatory or inhibitory postsynaptic potentials by depolarizing or hyperpolarizing, respectively, the postsynaptic membrane. In electrical synapses, the action potential is directly communicated to the postsynaptic cell through gap junctions—large channel proteins that connect the pre- and postsynaptic membranes. Synapses are not static structures and can be strengthened and weakened. Two mechanisms of synaptic plasticity are long-term potentiation and long-term depression.

## Art Connections

### Exercise:

#### Problem:

[\[link\]](#) Potassium channel blockers, such as amiodarone and procainamide, which are used to treat abnormal electrical activity in the heart, called cardiac dysrhythmia, impede the movement of  $K^+$  through voltage-gated  $K^+$  channels. Which part of the action potential would you expect potassium channels to affect?

---

#### Solution:



[\[link\]](#) Potassium channel blockers slow the repolarization phase, but have no effect on depolarization.

## Review Questions

### Exercise:

#### Problem:

For a neuron to fire an action potential, its membrane must reach \_\_\_\_\_.

- a. hyperpolarization
- b. the threshold of excitation
- c. the refractory period
- d. inhibitory postsynaptic potential

---

#### Solution:

B

### Exercise:

#### Problem:

After an action potential, the opening of additional voltage-gated \_\_\_\_\_ channels and the inactivation of sodium channels, cause the membrane to return to its resting membrane potential.

- a. sodium
- b. potassium
- c. calcium
- d. chloride

---

#### Solution:

B

**Exercise:****Problem:**

What is the term for protein channels that connect two neurons at an electrical synapse?

- a. synaptic vesicles
- b. voltage-gated ion channels
- c. gap junction protein
- d. sodium-potassium exchange pumps

---

**Solution:**

C

**Free Response****Exercise:****Problem:**

How does myelin aid propagation of an action potential along an axon? How do the nodes of Ranvier help this process?

---

**Solution:**

Myelin prevents the leak of current from the axon. Nodes of Ranvier allow the action potential to be regenerated at specific points along the axon. They also save energy for the cell since voltage-gated ion channels and sodium-potassium transporters are not needed along myelinated portions of the axon.

**Exercise:**

**Problem:** What are the main steps in chemical neurotransmission?

---

**Solution:**

An action potential travels along an axon until it depolarizes the membrane at an axon terminal. Depolarization of the membrane causes voltage-gated  $\text{Ca}^{2+}$  channels to open and  $\text{Ca}^{2+}$  to enter the cell. The intracellular calcium influx causes synaptic vesicles containing neurotransmitter to fuse with the presynaptic membrane. The neurotransmitter diffuses across the synaptic cleft and binds to receptors on the postsynaptic membrane. Depending on the specific neurotransmitter and postsynaptic receptor, this action can cause positive (excitatory postsynaptic potential) or negative (inhibitory postsynaptic potential) ions to enter the cell.

**Glossary**

action potential

self-propagating momentary change in the electrical potential of a neuron (or muscle) membrane

depolarization

change in the membrane potential to a less negative value

excitatory postsynaptic potential (EPSP)

depolarization of a postsynaptic membrane caused by neurotransmitter molecules released from a presynaptic cell

hyperpolarization

change in the membrane potential to a more negative value

inhibitory postsynaptic potential (IPSP)

hyperpolarization of a postsynaptic membrane caused by neurotransmitter molecules released from a presynaptic cell

long-term depression (LTD)

prolonged decrease in synaptic coupling between a pre- and postsynaptic cell

long-term potentiation (LTP)

prolonged increase in synaptic coupling between a pre-and postsynaptic cell

membrane potential

difference in electrical potential between the inside and outside of a cell

refractory period

period after an action potential when it is more difficult or impossible for an action potential to be fired; caused by inactivation of sodium channels and activation of additional potassium channels of the membrane

saltatory conduction

“jumping” of an action potential along an axon from one node of Ranvier to the next

summation

process of multiple presynaptic inputs creating EPSPs around the same time for the postsynaptic neuron to be sufficiently depolarized to fire an action potential

synaptic cleft

space between the presynaptic and postsynaptic membranes

synaptic vesicle

spherical structure that contains a neurotransmitter

threshold of excitation

level of depolarization needed for an action potential to fire

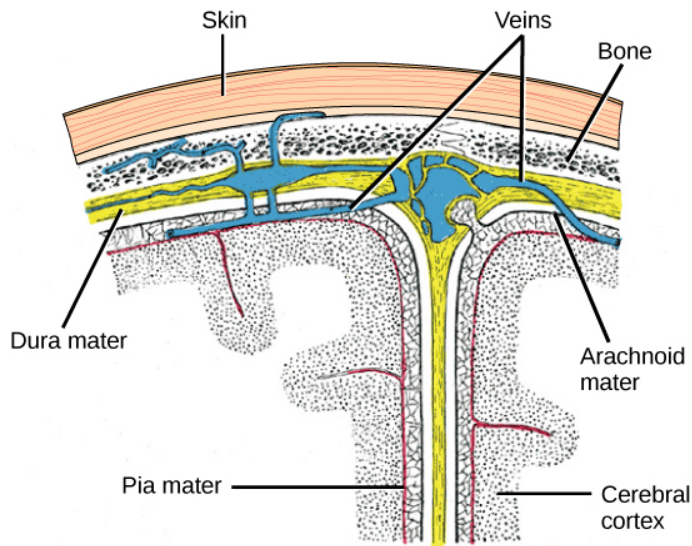
## The Central Nervous System

By the end of this section, you will be able to:

- Identify the spinal cord, cerebral lobes, and other brain areas on a diagram of the brain
- Describe the basic functions of the spinal cord, cerebral lobes, and other brain areas

The central nervous system (CNS) is made up of the brain, a part of which is shown in [\[link\]](#) and spinal cord and is covered with three layers of protective coverings called **meninges** (from the Greek word for membrane). The outermost layer is the **dura mater** (Latin for “hard mother”). As the Latin suggests, the primary function for this thick layer is to protect the brain and spinal cord. The dura mater also contains vein-like structures that carry blood from the brain back to the heart. The middle layer is the web-like **arachnoid mater**. The last layer is the **pia mater** (Latin for “soft mother”), which directly contacts and covers the brain and spinal cord like plastic wrap. The space between the arachnoid and pia mater is filled with **cerebrospinal fluid (CSF)**. CSF is produced by a tissue called **choroid plexus** in fluid-filled compartments in the CNS called **ventricles**. The brain floats in CSF, which acts as a cushion and shock absorber and makes the brain neutrally buoyant. CSF also functions to circulate chemical substances throughout the brain and into the spinal cord.

The entire brain contains only about 8.5 tablespoons of CSF, but CSF is constantly produced in the ventricles. This creates a problem when a ventricle is blocked—the CSF builds up and creates swelling and the brain is pushed against the skull. This swelling condition is called hydrocephalus (“water head”) and can cause seizures, cognitive problems, and even death if a shunt is not inserted to remove the fluid and pressure.



The cerebral cortex is covered by three layers of meninges: the dura, arachnoid, and pia maters. (credit: modification of work by Gray's Anatomy)

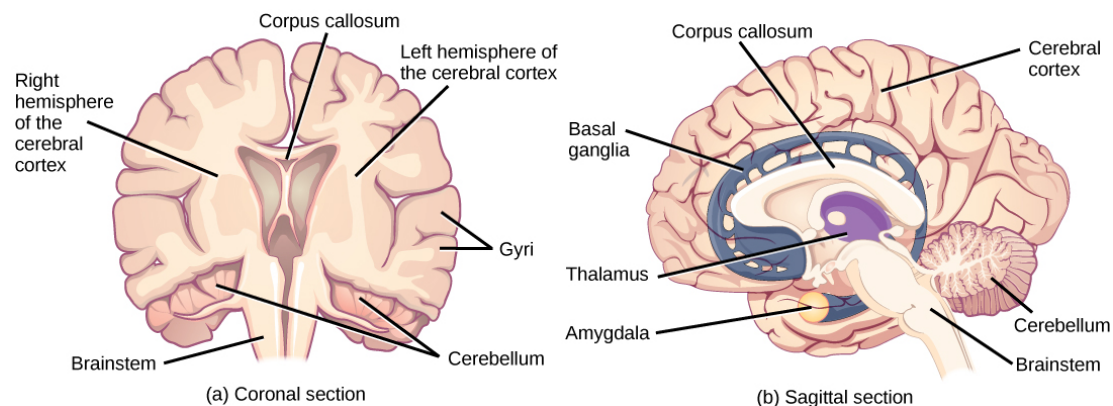
## Brain

The brain is the part of the central nervous system that is contained in the cranial cavity of the skull. It includes the cerebral cortex, limbic system, basal ganglia, thalamus, hypothalamus, and cerebellum. There are three different ways that a brain can be sectioned in order to view internal structures: a sagittal section cuts the brain left to right, as shown in [\[link\]](#)**b**, a coronal section cuts the brain front to back, as shown in [\[link\]](#)**a**, and a horizontal section cuts the brain top to bottom.

## Cerebral Cortex

The outermost part of the brain is a thick piece of nervous system tissue called the **cerebral cortex**, which is folded into hills called **gyri** (singular:

gyrus) and valleys called **sulci** (singular: sulcus). The cortex is made up of two hemispheres—right and left—which are separated by a large sulcus. A thick fiber bundle called the **corpus callosum** (Latin: “tough body”) connects the two hemispheres and allows information to be passed from one side to the other. Although there are some brain functions that are localized more to one hemisphere than the other, the functions of the two hemispheres are largely redundant. In fact, sometimes (very rarely) an entire hemisphere is removed to treat severe epilepsy. While patients do suffer some deficits following the surgery, they can have surprisingly few problems, especially when the surgery is performed on children who have very immature nervous systems.



These illustrations show the (a) coronal and (b) sagittal sections of the human brain.

In other surgeries to treat severe epilepsy, the corpus callosum is cut instead of removing an entire hemisphere. This causes a condition called split-brain, which gives insights into unique functions of the two hemispheres. For example, when an object is presented to patients' left visual field, they may be unable to verbally name the object (and may claim to not have seen an object at all). This is because the visual input from the left visual field crosses and enters the right hemisphere and cannot then signal to the speech center, which generally is found in the left side of the brain. Remarkably, if a split-brain patient is asked to pick up a specific object out of a group of

objects with the left hand, the patient will be able to do so but will still be unable to vocally identify it.

**Note:**

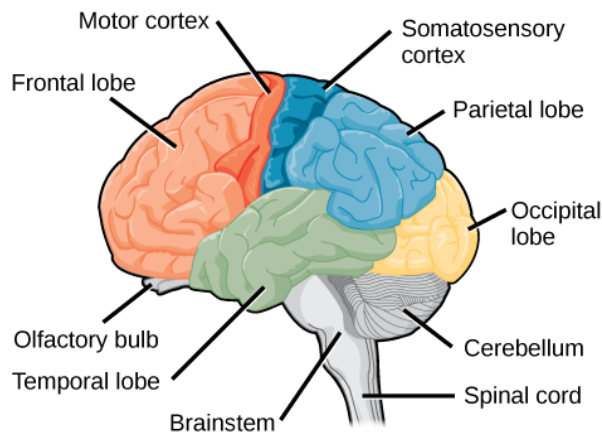
Link to Learning



See [this website](#) to learn more about split-brain patients and to play a game where you can model the split-brain experiments yourself.

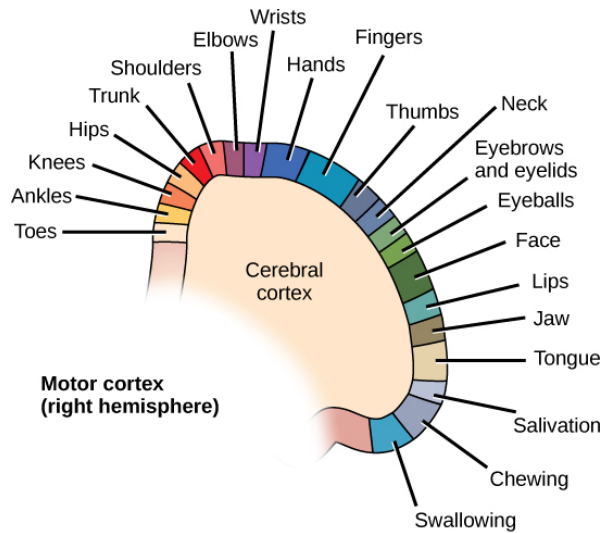
Each cortical hemisphere contains regions called lobes that are involved in different functions. Scientists use various techniques to determine what brain areas are involved in different functions: they examine patients who have had injuries or diseases that affect specific areas and see how those areas are related to functional deficits. They also conduct animal studies where they stimulate brain areas and see if there are any behavioral changes. They use a technique called transcranial magnetic stimulation (TMS) to temporarily deactivate specific parts of the cortex using strong magnets placed outside the head; and they use functional magnetic resonance imaging (fMRI) to look at changes in oxygenated blood flow in particular brain regions that correlate with specific behavioral tasks. These techniques, and others, have given great insight into the functions of different brain regions but have also showed that any given brain area can be involved in more than one behavior or process, and any given behavior or process generally involves neurons in multiple brain areas. That being said, each hemisphere of the mammalian cerebral cortex can be broken down into four functionally and spatially defined lobes: frontal, parietal, temporal, and occipital. [\[link\]](#) illustrates these four lobes of the human cerebral cortex.





The human cerebral cortex includes the frontal, parietal, temporal, and occipital lobes.

The **frontal lobe** is located at the front of the brain, over the eyes. This lobe contains the olfactory bulb, which processes smells. The frontal lobe also contains the motor cortex, which is important for planning and implementing movement. Areas within the motor cortex map to different muscle groups, and there is some organization to this map, as shown in [\[link\]](#). For example, the neurons that control movement of the fingers are next to the neurons that control movement of the hand. Neurons in the frontal lobe also control cognitive functions like maintaining attention, speech, and decision-making. Studies of humans who have damaged their frontal lobes show that parts of this area are involved in personality, socialization, and assessing risk.



Different parts of the motor cortex control different muscle groups. Muscle groups that are neighbors in the body are generally controlled by neighboring regions of the motor cortex as well. For example, the neurons that control finger movement are near the neurons that control hand movement.

The **parietal lobe** is located at the top of the brain. Neurons in the parietal lobe are involved in speech and also reading. Two of the parietal lobe's main functions are processing **somatosensation**—touch sensations like pressure, pain, heat, cold—and processing **proprioception**—the sense of how parts of the body are oriented in space. The parietal lobe contains a somatosensory map of the body similar to the motor cortex.

The **occipital lobe** is located at the back of the brain. It is primarily involved in vision—seeing, recognizing, and identifying the visual world.

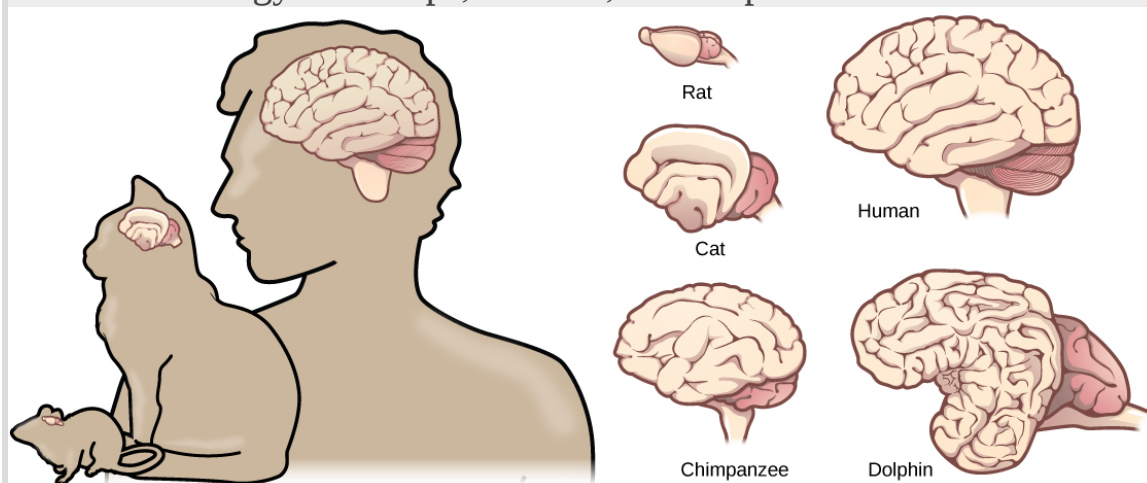
The **temporal lobe** is located at the base of the brain by your ears and is primarily involved in processing and interpreting sounds. It also contains the **hippocampus** (Greek for “seahorse”)—a structure that processes memory formation. The hippocampus is illustrated in [\[link\]](#). The role of the hippocampus in memory was partially determined by studying one famous epileptic patient, HM, who had both sides of his hippocampus removed in an attempt to cure his epilepsy. His seizures went away, but he could no longer form new memories (although he could remember some facts from before his surgery and could learn new motor tasks).

### Note:

#### Evolution Connection

#### Cerebral Cortex

Compared to other vertebrates, mammals have exceptionally large brains for their body size. An entire alligator’s brain, for example, would fill about one and a half teaspoons. This increase in brain to body size ratio is especially pronounced in apes, whales, and dolphins. While this increase in overall brain size doubtlessly played a role in the evolution of complex behaviors unique to mammals, it does not tell the whole story. Scientists have found a relationship between the relatively high surface area of the cortex and the intelligence and complex social behaviors exhibited by some mammals. This increased surface area is due, in part, to increased folding of the cortical sheet (more sulci and gyri). For example, a rat cortex is very smooth with very few sulci and gyri. Cat and sheep cortices have more sulci and gyri. Chimps, humans, and dolphins have even more.



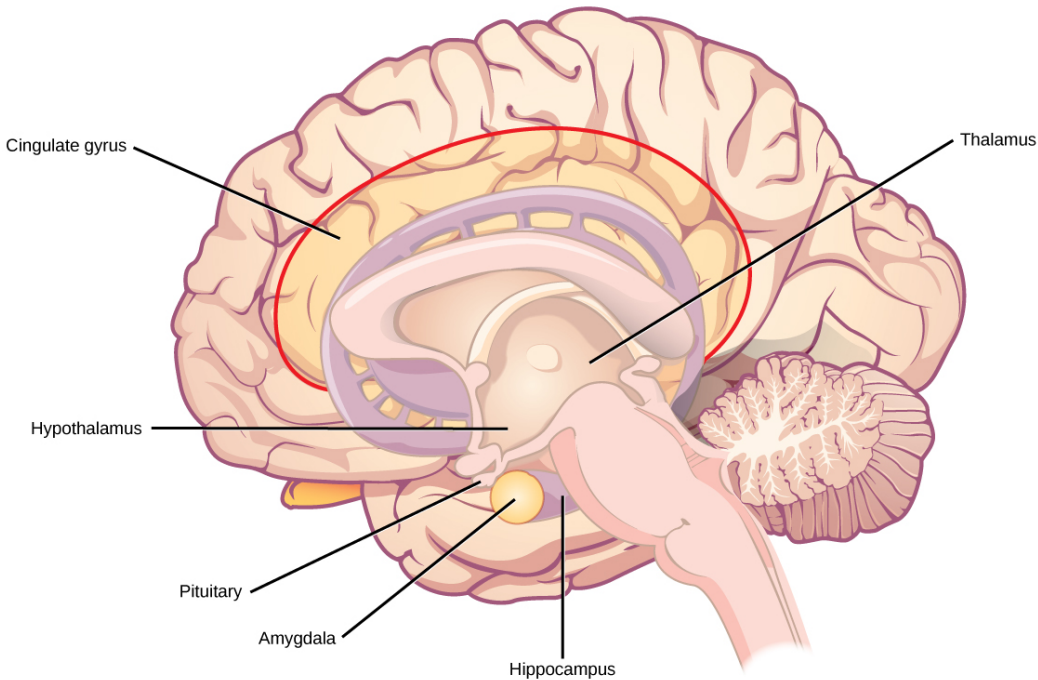
Mammals have larger brain-to-body ratios than other vertebrates. Within mammals, increased cortical folding and surface area is correlated with complex behavior.

## Basal Ganglia

Interconnected brain areas called the **basal ganglia** (or **basal nuclei**), shown in [\[link\]](#)**b**, play important roles in movement control and posture. Damage to the basal ganglia, as in Parkinson’s disease, leads to motor impairments like a shuffling gait when walking. The basal ganglia also regulate motivation. For example, when a wasp sting led to bilateral basal ganglia damage in a 25-year-old businessman, he began to spend all his days in bed and showed no interest in anything or anybody. But when he was externally stimulated—as when someone asked to play a card game with him—he was able to function normally. Interestingly, he and other similar patients do not report feeling bored or frustrated by their state.

## Thalamus

The **thalamus** (Greek for “inner chamber”), illustrated in [\[link\]](#), acts as a gateway to and from the cortex. It receives sensory and motor inputs from the body and also receives feedback from the cortex. This feedback mechanism can modulate conscious awareness of sensory and motor inputs depending on the attention and arousal state of the animal. The thalamus helps regulate consciousness, arousal, and sleep states. A rare genetic disorder called fatal familial insomnia causes the degeneration of thalamic neurons and glia. This disorder prevents affected patients from being able to sleep, among other symptoms, and is eventually fatal.



The limbic system regulates emotion and other behaviors. It includes parts of the cerebral cortex located near the center of the brain, including the cingulate gyrus and the hippocampus as well as the thalamus, hypothalamus and amygdala.

## Hypothalamus

Below the thalamus is the **hypothalamus**, shown in [\[link\]](#). The hypothalamus controls the endocrine system by sending signals to the pituitary gland, a pea-sized endocrine gland that releases several different hormones that affect other glands as well as other cells. This relationship means that the hypothalamus regulates important behaviors that are controlled by these hormones. The hypothalamus is the body's thermostat—it makes sure key functions like food and water intake, energy expenditure, and body temperature are kept at appropriate levels. Neurons within the

hypothalamus also regulate circadian rhythms, sometimes called sleep cycles.

## **Limbic System**

The **limbic system** is a connected set of structures that regulates emotion, as well as behaviors related to fear and motivation. It plays a role in memory formation and includes parts of the thalamus and hypothalamus as well as the hippocampus. One important structure within the limbic system is a temporal lobe structure called the **amygdala** (Greek for “almond”), illustrated in [\[link\]](#). The two amygdala are important both for the sensation of fear and for recognizing fearful faces. The **cingulate gyrus** helps regulate emotions and pain.

## **Cerebellum**

The **cerebellum** (Latin for “little brain”), shown in [\[link\]](#), sits at the base of the brain on top of the brainstem. The cerebellum controls balance and aids in coordinating movement and learning new motor tasks.

## **Brainstem**

The **brainstem**, illustrated in [\[link\]](#), connects the rest of the brain with the spinal cord. It consists of the midbrain, medulla oblongata, and the pons. Motor and sensory neurons extend through the brainstem allowing for the relay of signals between the brain and spinal cord. Ascending neural pathways cross in this section of the brain allowing the left hemisphere of the cerebrum to control the right side of the body and vice versa. The brainstem coordinates motor control signals sent from the brain to the body. The brainstem controls several important functions of the body including alertness, arousal, breathing, blood pressure, digestion, heart rate, swallowing, walking, and sensory and motor information integration.

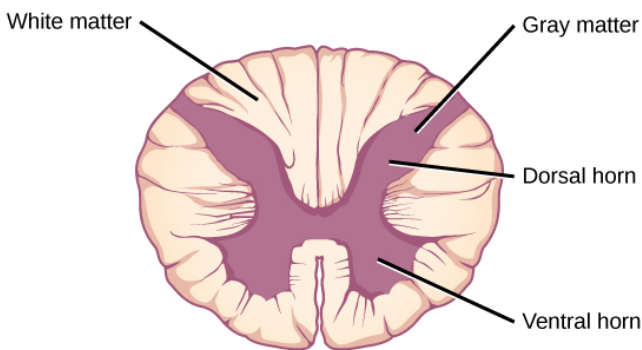
## Spinal Cord

Connecting to the brainstem and extending down the body through the spinal column is the **spinal cord**, shown in [\[link\]](#). The spinal cord is a thick bundle of nerve tissue that carries information about the body to the brain and from the brain to the body. The spinal cord is contained within the bones of the vertebrate column but is able to communicate signals to and from the body through its connections with spinal nerves (part of the peripheral nervous system). A cross-section of the spinal cord looks like a white oval containing a gray butterfly-shape, as illustrated in [\[link\]](#). Myelinated axons make up the “white matter” and neuron and glial cell bodies make up the “gray matter.” Gray matter is also composed of interneurons, which connect two neurons each located in different parts of the body. Axons and cell bodies in the dorsal (facing the back of the animal) spinal cord convey mostly sensory information from the body to the brain. Axons and cell bodies in the ventral (facing the front of the animal) spinal cord primarily transmit signals controlling movement from the brain to the body.

The spinal cord also controls motor reflexes. These reflexes are quick, unconscious movements—like automatically removing a hand from a hot object. Reflexes are so fast because they involve local synaptic connections. For example, the knee reflex that a doctor tests during a routine physical is controlled by a single synapse between a sensory neuron and a motor neuron. While a reflex may only require the involvement of one or two synapses, synapses with interneurons in the spinal column transmit information to the brain to convey what happened (the knee jerked, or the hand was hot).

In the United States, there around 10,000 spinal cord injuries each year. Because the spinal cord is the information superhighway connecting the brain with the body, damage to the spinal cord can lead to paralysis. The extent of the paralysis depends on the location of the injury along the spinal cord and whether the spinal cord was completely severed. For example, if the spinal cord is damaged at the level of the neck, it can cause paralysis from the neck down, whereas damage to the spinal column further down may limit paralysis to the legs. Spinal cord injuries are notoriously difficult

to treat because spinal nerves do not regenerate, although ongoing research suggests that stem cell transplants may be able to act as a bridge to reconnect severed nerves. Researchers are also looking at ways to prevent the inflammation that worsens nerve damage after injury. One such treatment is to pump the body with cold saline to induce hypothermia. This cooling can prevent swelling and other processes that are thought to worsen spinal cord injuries.



A cross-section of the spinal cord shows gray matter (containing cell bodies and interneurons) and white matter (containing axons).

## Section Summary

The vertebrate central nervous system contains the brain and the spinal cord, which are covered and protected by three meninges. The brain contains structurally and functionally defined regions. In mammals, these include the cortex (which can be broken down into four primary functional lobes: frontal, temporal, occipital, and parietal), basal ganglia, thalamus, hypothalamus, limbic system, cerebellum, and brainstem—although structures in some of these designations overlap. While functions may be primarily localized to one structure in the brain, most complex functions, like language and sleep, involve neurons in multiple brain regions. The



spinal cord is the information superhighway that connects the brain with the rest of the body through its connections with peripheral nerves. It transmits sensory and motor input and also controls motor reflexes.

## Review Questions

### Exercise:

**Problem:** The \_\_\_\_\_ lobe contains the visual cortex.

- a. frontal
- b. parietal
- c. temporal
- d. occipital

---

### Solution:

D

### Exercise:

**Problem:** The \_\_\_\_\_ connects the two cerebral hemispheres.

- a. limbic system
- b. corpus callosum
- c. cerebellum
- d. pituitary

---

### Solution:

B

### Exercise:

**Problem:** Neurons in the \_\_\_\_\_ control motor reflexes.

- a. thalamus
- b. spinal cord
- c. parietal lobe
- d. hippocampus

---

**Solution:**

B

## Free Response

**Exercise:**

**Problem:**

What methods can be used to determine the function of a particular brain region?

---

**Solution:**

To determine the function of a specific brain area, scientists can look at patients who have damage in that brain area and see what symptoms they exhibit. Researchers can disable the brain structure temporarily using transcranial magnetic stimulation. They can disable or remove the area in an animal model. fMRI can be used to correlate specific functions with increased blood flow to brain regions.

**Exercise:**

**Problem:**What are the main functions of the spinal cord?

---

**Solution:**

The spinal cord transmits sensory information from the body to the brain and motor commands from the brain to the body through its connections with peripheral nerves. It also controls motor reflexes.

## **Glossary**

amygdala

structure within the limbic system that processes fear

arachnoid mater

spiderweb-like middle layer of the meninges that cover the central nervous system

basal ganglia

interconnected collections of cells in the brain that are involved in movement and motivation; also known as basal nuclei

basal nuclei

see basal ganglia

brainstem

portion of the brain that connects with the spinal cord; controls basic nervous system functions like breathing, heart rate, and swallowing

cerebellum

brain structure involved in posture, motor coordination, and learning new motor actions

cerebral cortex

outermost sheet of brain tissue; involved in many higher-order functions

choroid plexus

spongy tissue within ventricles that produces cerebrospinal fluid

cingulate gyrus

helps regulate emotions and pain; thought to directly drive the body's conscious response to unpleasant experiences

corpus callosum

thick fiber bundle that connects the cerebral hemispheres

cerebrospinal fluid (CSF)

clear liquid that surrounds the brain and spinal cord and fills the ventricles and central canal; acts as a shock absorber and circulates material throughout the brain and spinal cord.

dura mater

tough outermost layer that covers the central nervous system

frontal lobe

part of the cerebral cortex that contains the motor cortex and areas involved in planning, attention, and language

gyrus

(plural: gyri) ridged protrusions in the cortex

hippocampus

brain structure in the temporal lobe involved in processing memories

hypothalamus

brain structure that controls hormone release and body homeostasis

limbic system

connected brain areas that process emotion and motivation

meninge

membrane that covers and protects the central nervous system

occipital lobe

part of the cerebral cortex that contains visual cortex and processes visual stimuli

parietal lobe

part of the cerebral cortex involved in processing touch and the sense of the body in space

pia mater

thin membrane layer directly covering the brain and spinal cord

proprioception

sense about how parts of the body are oriented in space

somatosensation

sense of touch

spinal cord

thick fiber bundle that connects the brain with peripheral nerves;  
transmits sensory and motor information; contains neurons that control  
motor reflexes

sulcus

(plural: sulci) indents or “valleys” in the cortex

temporal lobe

part of the cerebral cortex that processes auditory input; parts of the  
temporal lobe are involved in speech, memory, and emotion processing

thalamus

brain area that relays sensory information to the cortex

ventricle

cavity within brain that contains cerebrospinal fluid

## The Peripheral Nervous System

By the end of this section, you will be able to:

- Describe the organization and functions of the sympathetic and parasympathetic nervous systems
- Describe the organization and function of the sensory-somatic nervous system

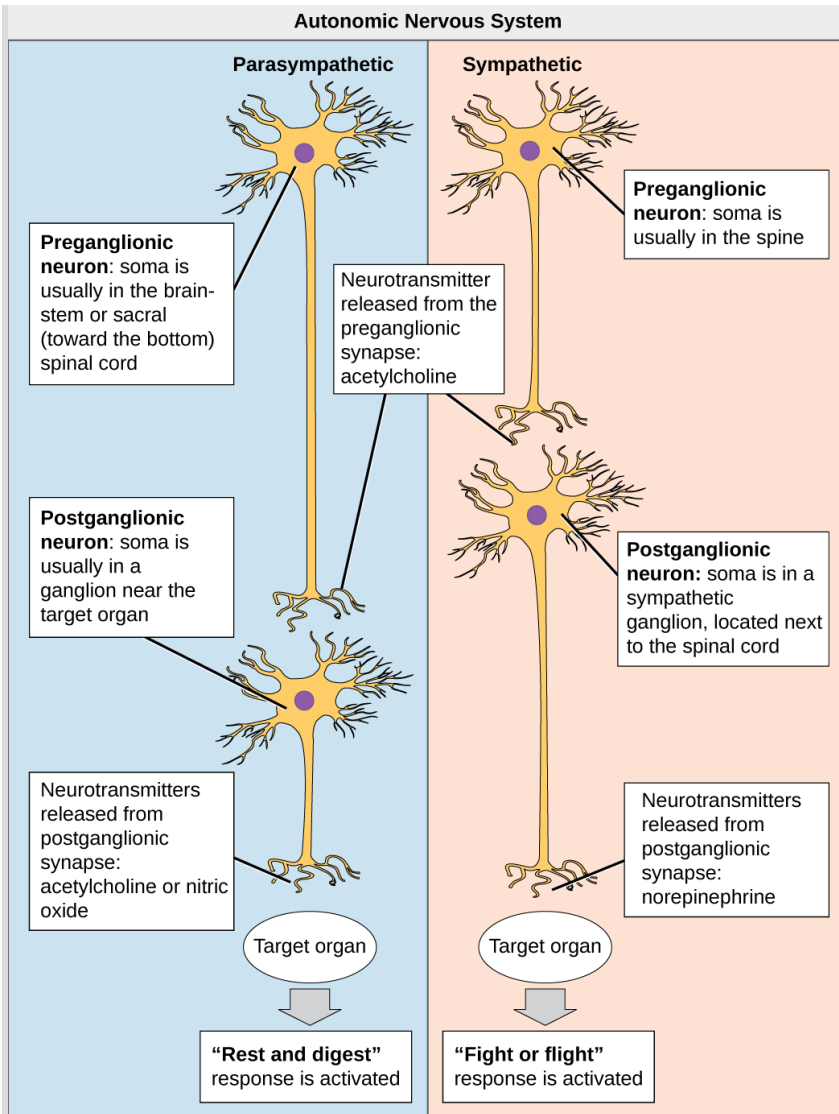
The peripheral nervous system (PNS) is the connection between the central nervous system and the rest of the body. The CNS is like the power plant of the nervous system. It creates the signals that control the functions of the body. The PNS is like the wires that go to individual houses. Without those “wires,” the signals produced by the CNS could not control the body (and the CNS would not be able to receive sensory information from the body either).

The PNS can be broken down into the **autonomic nervous system**, which controls bodily functions without conscious control, and the **sensory-somatic nervous system**, which transmits sensory information from the skin, muscles, and sensory organs to the CNS and sends motor commands from the CNS to the muscles.

## Autonomic Nervous System

### Note:

Art Connection



In the autonomic nervous system, a preganglionic neuron of the CNS synapses with a postganglionic neuron of the PNS. The postganglionic neuron, in turn, acts on a target organ. Autonomic responses are mediated by the sympathetic and the parasympathetic systems, which are antagonistic to one another. The sympathetic system activates the "fight or flight" response, while the parasympathetic system activates the "rest and digest" response.

Which of the following statements is false?

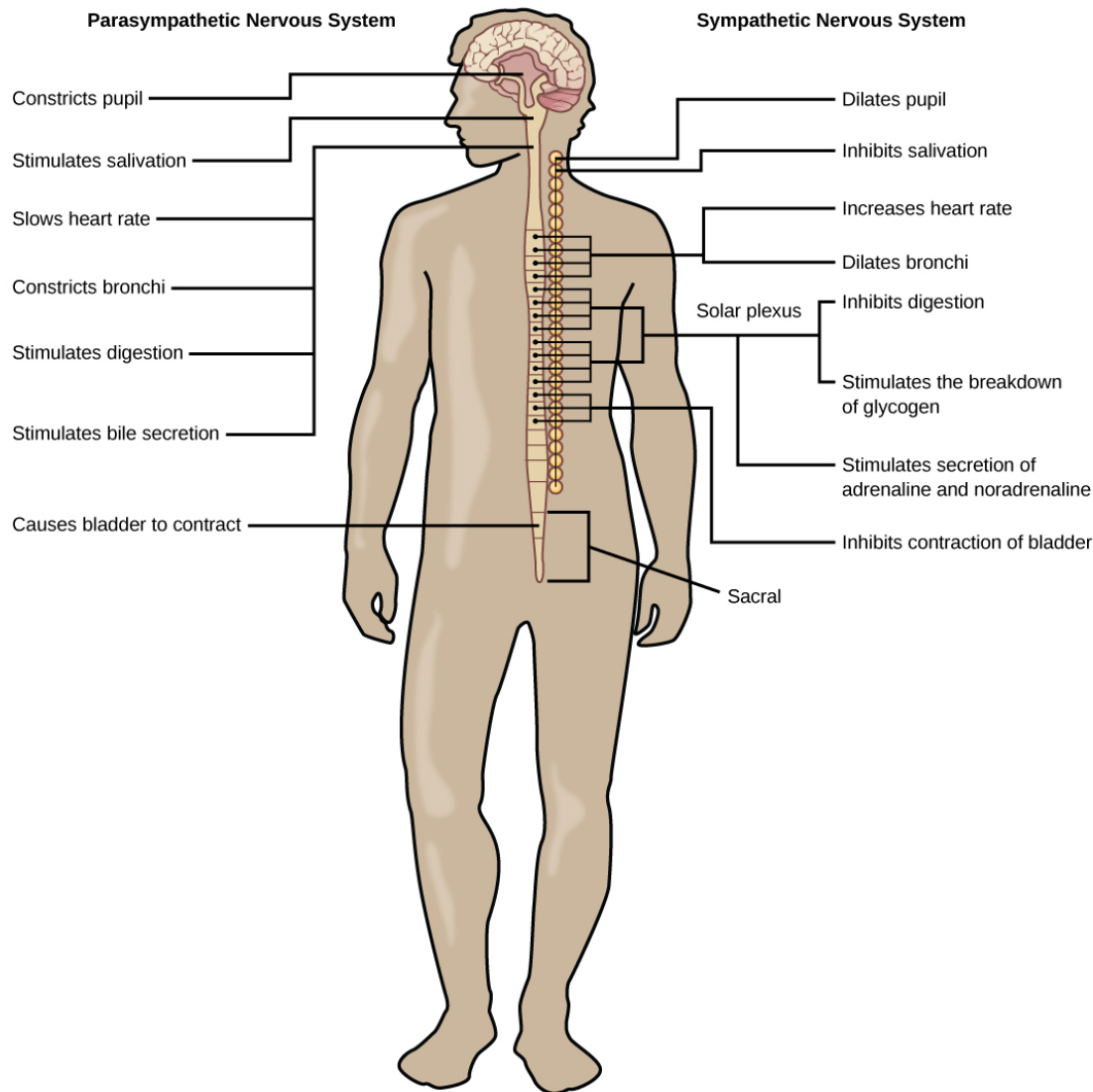
- a. The parasympathetic pathway is responsible for resting the body, while the sympathetic pathway is responsible for preparing for an emergency.
- b. Most preganglionic neurons in the sympathetic pathway originate in the spinal cord.
- c. Slowing of the heartbeat is a parasympathetic response.
- d. Parasympathetic neurons are responsible for releasing norepinephrine on the target organ, while sympathetic neurons are responsible for releasing acetylcholine.

The autonomic nervous system serves as the relay between the CNS and the internal organs. It controls the lungs, the heart, smooth muscle, and exocrine and endocrine glands. The autonomic nervous system controls these organs largely without conscious control; it can continuously monitor the conditions of these different systems and implement changes as needed. Signaling to the target tissue usually involves two synapses: a preganglionic neuron (originating in the CNS) synapses to a neuron in a ganglion that, in turn, synapses on the target organ, as illustrated in [\[link\]](#). There are two divisions of the autonomic nervous system that often have opposing effects: the sympathetic nervous system and the parasympathetic nervous system.

## Sympathetic Nervous System

The **sympathetic nervous system** is responsible for the “fight or flight” response that occurs when an animal encounters a dangerous situation. One way to remember this is to think of the surprise a person feels when encountering a snake (“snake” and “sympathetic” both begin with “s”). Examples of functions controlled by the sympathetic nervous system include an accelerated heart rate and inhibited digestion. These functions help prepare an organism’s body for the physical strain required to escape a potentially dangerous situation or to fend off a predator.





The sympathetic and parasympathetic nervous systems often have opposing effects on target organs.

Most preganglionic neurons in the sympathetic nervous system originate in the spinal cord, as illustrated in [\[link\]](#). The axons of these neurons release **acetylcholine** on postganglionic neurons within sympathetic ganglia (the sympathetic ganglia form a chain that extends alongside the spinal cord). The acetylcholine activates the postganglionic neurons. Postganglionic neurons then release **norepinephrine** onto target organs. As anyone who has ever felt a rush before a big test, speech, or athletic event can attest, the

effects of the sympathetic nervous system are quite pervasive. This is both because one preganglionic neuron synapses on multiple postganglionic neurons, amplifying the effect of the original synapse, and because the adrenal gland also releases norepinephrine (and the closely related hormone epinephrine) into the blood stream. The physiological effects of this norepinephrine release include dilating the trachea and bronchi (making it easier for the animal to breathe), increasing heart rate, and moving blood from the skin to the heart, muscles, and brain (so the animal can think and run). The strength and speed of the sympathetic response helps an organism avoid danger, and scientists have found evidence that it may also increase LTP—allowing the animal to remember the dangerous situation and avoid it in the future.

## **Parasympathetic Nervous System**

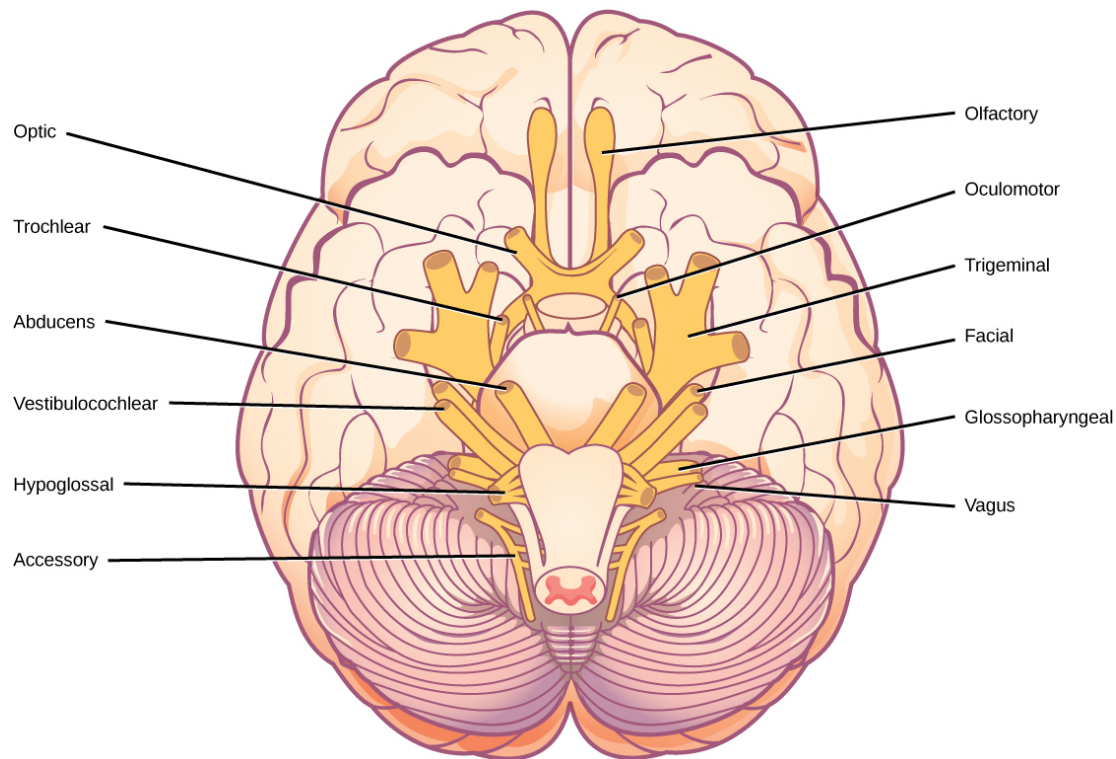
While the sympathetic nervous system is activated in stressful situations, the **parasympathetic nervous system** allows an animal to “rest and digest.” One way to remember this is to think that during a restful situation like a picnic, the parasympathetic nervous system is in control (“picnic” and “parasympathetic” both start with “p”). Parasympathetic preganglionic neurons have cell bodies located in the brainstem and in the sacral (toward the bottom) spinal cord, as shown in [\[link\]](#). The axons of the preganglionic neurons release acetylcholine on the postganglionic neurons, which are generally located very near the target organs. Most postganglionic neurons release acetylcholine onto target organs, although some release nitric oxide.

The parasympathetic nervous system resets organ function after the sympathetic nervous system is activated (the common adrenaline dump you feel after a ‘fight-or-flight’ event). Effects of acetylcholine release on target organs include slowing of heart rate, lowered blood pressure, and stimulation of digestion.

## **Sensory-Somatic Nervous System**

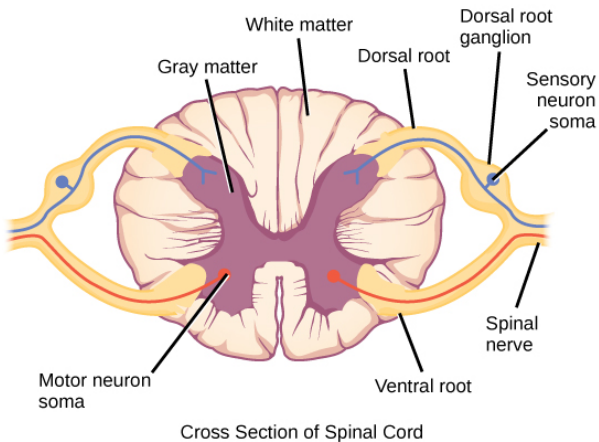
The sensory-somatic nervous system is made up of cranial and spinal nerves and contains both sensory and motor neurons. Sensory neurons transmit sensory information from the skin, skeletal muscle, and sensory organs to the CNS. Motor neurons transmit messages about desired movement from the CNS to the muscles to make them contract. Without its sensory-somatic nervous system, an animal would be unable to process any information about its environment (what it sees, feels, hears, and so on) and could not control motor movements. Unlike the autonomic nervous system, which has two synapses between the CNS and the target organ, sensory and motor neurons have only one synapse—one ending of the neuron is at the organ and the other directly contacts a CNS neuron. Acetylcholine is the main neurotransmitter released at these synapses.

Humans have 12 **cranial nerves**, nerves that emerge from or enter the skull (cranium), as opposed to the spinal nerves, which emerge from the vertebral column. Each cranial nerve is accorded a name, which are detailed in [\[link\]](#). Some cranial nerves transmit only sensory information. For example, the olfactory nerve transmits information about smells from the nose to the brainstem. Other cranial nerves transmit almost solely motor information. For example, the oculomotor nerve controls the opening and closing of the eyelid and some eye movements. Other cranial nerves contain a mix of sensory and motor fibers. For example, the glossopharyngeal nerve has a role in both taste (sensory) and swallowing (motor).



The human brain contains 12 cranial nerves that receive sensory input and control motor output for the head and neck.

**Spinal nerves** transmit sensory and motor information between the spinal cord and the rest of the body. Each of the 31 spinal nerves (in humans) contains both sensory and motor axons. The sensory neuron cell bodies are grouped in structures called dorsal root ganglia and are shown in [\[link\]](#). Each sensory neuron has one projection—with a sensory receptor ending in skin, muscle, or sensory organs—and another that synapses with a neuron in the dorsal spinal cord. Motor neurons have cell bodies in the ventral gray matter of the spinal cord that project to muscle through the ventral root. These neurons are usually stimulated by interneurons within the spinal cord but are sometimes directly stimulated by sensory neurons.



Spinal nerves contain both sensory and motor axons. The somas of sensory neurons are located in dorsal root ganglia. The somas of motor neurons are found in the ventral portion of the gray matter of the spinal cord.

## Section Summary

The peripheral nervous system contains both the autonomic and sensory-somatic nervous systems. The autonomic nervous system provides unconscious control over visceral functions and has two divisions: the sympathetic and parasympathetic nervous systems. The sympathetic nervous system is activated in stressful situations to prepare the animal for a “fight or flight” response. The parasympathetic nervous system is active during restful periods. The sensory-somatic nervous system is made of cranial and spinal nerves that transmit sensory information from skin and muscle to the CNS and motor commands from the CNS to the muscles.

## Art Connections

### Exercise:

**Problem:** [\[link\]](#) Which of the following statements is false?

- a. The parasympathetic pathway is responsible for relaxing the body, while the sympathetic pathway is responsible for preparing for an emergency.
  - b. Most preganglionic neurons in the sympathetic pathway originate in the spinal cord.
  - c. Slowing of the heartbeat is a parasympathetic response.
  - d. Parasympathetic neurons are responsible for releasing norepinephrine on the target organ, while sympathetic neurons are responsible for releasing acetylcholine.
- 

**Solution:**

[\[link\]](#) D

### Review Questions

#### Exercise:

**Problem:** Activation of the sympathetic nervous system causes:

- a. increased blood flow into the skin
  - b. a decreased heart rate
  - c. an increased heart rate
  - d. increased digestion
- 

**Solution:**

C

#### Exercise:

**Problem:**

Where are parasympathetic preganglionic cell bodies located?

- a. cerebellum
- b. brainstem
- c. dorsal root ganglia
- d. skin

---

**Solution:**

B

**Exercise:**

**Problem:**\_\_\_\_\_ is released by motor nerve endings onto muscle.

- a. Acetylcholine
- b. Norepinephrine
- c. Dopamine
- d. Serotonin

---

**Solution:**

A

**Free Response****Exercise:****Problem:**

What are the main differences between the sympathetic and parasympathetic branches of the autonomic nervous system?

---

**Solution:**

The sympathetic nervous system prepares the body for “fight or flight,” whereas the parasympathetic nervous system allows the body to “rest and digest.” Sympathetic neurons release norepinephrine onto target organs; parasympathetic neurons release acetylcholine. Sympathetic neuron cell bodies are located in sympathetic ganglia. Parasympathetic neuron cell bodies are located in the brainstem and sacral spinal cord. Activation of the sympathetic nervous system increases heart rate and blood pressure and decreases digestion and blood flow to the skin. Activation of the parasympathetic nervous system decreases heart rate and blood pressure and increases digestion and blood flow to the skin.

**Exercise:****Problem:**

What are the main functions of the sensory-somatic nervous system?

---

**Solution:**

The sensory-somatic nervous system transmits sensory information from the skin, muscles, and sensory organs to the CNS. It also sends motor commands from the CNS to the muscles, causing them to contract.

**Glossary****acetylcholine**

neurotransmitter released by neurons in the central nervous system and peripheral nervous system

**autonomic nervous system**

part of the peripheral nervous system that controls bodily functions

**cranial nerve**

sensory and/or motor nerve that emanates from the brain

**norepinephrine**



neurotransmitter and hormone released by activation of the sympathetic nervous system

parasympathetic nervous system

division of autonomic nervous system that regulates visceral functions during rest and digestion

sensory-somatic nervous system

system of sensory and motor nerves

spinal nerve

nerve projecting between skin or muscle and spinal cord

sympathetic nervous system

division of autonomic nervous system activated during stressful “fight or flight” situations

## Nervous System Disorders

By the end of this section, you will be able to:

- Describe the symptoms, potential causes, and treatment of several examples of nervous system disorders

A nervous system that functions correctly is a fantastically complex, well-oiled machine—synapses fire appropriately, muscles move when needed, memories are formed and stored, and emotions are well regulated.

Unfortunately, each year millions of people in the United States deal with some sort of nervous system disorder. While scientists have discovered potential causes of many of these diseases, and viable treatments for some, ongoing research seeks to find ways to better prevent and treat all of these disorders.

## Neurodegenerative Disorders

**Neurodegenerative disorders** are illnesses characterized by a loss of nervous system functioning that are usually caused by neuronal death. These diseases generally worsen over time as more and more neurons die. The symptoms of a particular neurodegenerative disease are related to where in the nervous system the death of neurons occurs. Spinocerebellar ataxia, for example, leads to neuronal death in the cerebellum. The death of these neurons causes problems in balance and walking. Neurodegenerative disorders include Huntington's disease, amyotrophic lateral sclerosis, Alzheimer's disease and other types of dementia disorders, and Parkinson's disease. Here, Alzheimer's and Parkinson's disease will be discussed in more depth.

### Alzheimer's Disease

**Alzheimer's disease** is the most common cause of dementia in the elderly. In 2012, an estimated 5.4 million Americans suffered from Alzheimer's disease, and payments for their care are estimated at \$200 billion. Roughly one in every eight people age 65 or older has the disease. Due to the aging

of the baby-boomer generation, there are projected to be as many as 13 million Alzheimer's patients in the United States in the year 2050.

Symptoms of Alzheimer's disease include disruptive memory loss, confusion about time or place, difficulty planning or executing tasks, poor judgment, and personality changes. Problems smelling certain scents can also be indicative of Alzheimer's disease and may serve as an early warning sign. Many of these symptoms are also common in people who are aging normally, so it is the severity and longevity of the symptoms that determine whether a person is suffering from Alzheimer's.

Alzheimer's disease was named for Alois Alzheimer, a German psychiatrist who published a report in 1911 about a woman who showed severe dementia symptoms. Along with his colleagues, he examined the woman's brain following her death and reported the presence of abnormal clumps, which are now called amyloid plaques, along with tangled brain fibers called neurofibrillary tangles. Amyloid plaques, neurofibrillary tangles, and an overall shrinking of brain volume are commonly seen in the brains of Alzheimer's patients. Loss of neurons in the hippocampus is especially severe in advanced Alzheimer's patients. [\[link\]](#) compares a normal brain to the brain of an Alzheimer's patient. Many research groups are examining the causes of these hallmarks of the disease.

One form of the disease is usually caused by mutations in one of three known genes. This rare form of early onset Alzheimer's disease affects fewer than five percent of patients with the disease and causes dementia beginning between the ages of 30 and 60. The more prevalent, late-onset form of the disease likely also has a genetic component. One particular gene, apolipoprotein E (APOE) has a variant (E4) that increases a carrier's likelihood of getting the disease. Many other genes have been identified that might be involved in the pathology.

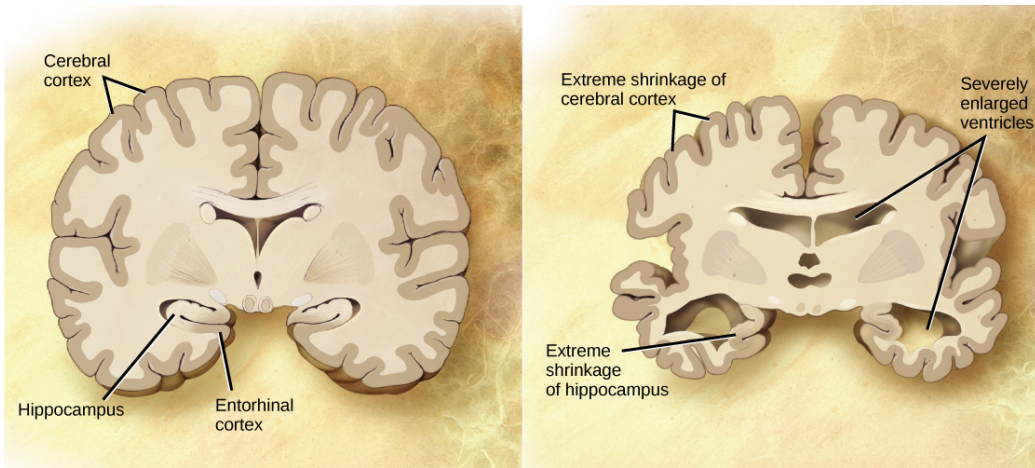
**Note:**

Link to Learning



Visit [this website](#) for video links discussing genetics and Alzheimer's disease.

Unfortunately, there is no cure for Alzheimer's disease. Current treatments focus on managing the symptoms of the disease. Because decrease in the activity of cholinergic neurons (neurons that use the neurotransmitter acetylcholine) is common in Alzheimer's disease, several drugs used to treat the disease work by increasing acetylcholine neurotransmission, often by inhibiting the enzyme that breaks down acetylcholine in the synaptic cleft. Other clinical interventions focus on behavioral therapies like psychotherapy, sensory therapy, and cognitive exercises. Since Alzheimer's disease appears to hijack the normal aging process, research into prevention is prevalent. Smoking, obesity, and cardiovascular problems may be risk factors for the disease, so treatments for those may also help to prevent Alzheimer's disease. Some studies have shown that people who remain intellectually active by playing games, reading, playing musical instruments, and being socially active in later life have a reduced risk of developing the disease.



Compared to a normal brain (left), the brain from a patient with Alzheimer's disease (right) shows a dramatic neurodegeneration, particularly within the ventricles and hippocampus. (credit: modification of work by "Garrando"/Wikimedia Commons based on original images by ADEAR: "Alzheimer's Disease Education and Referral Center, a service of the National Institute on Aging")

## Parkinson's Disease

Like Alzheimer's disease, **Parkinson's disease** is a neurodegenerative disease. It was first characterized by James Parkinson in 1817. Each year, 50,000-60,000 people in the United States are diagnosed with the disease. Parkinson's disease causes the loss of dopamine neurons in the substantia nigra, a midbrain structure that regulates movement. Loss of these neurons causes many symptoms including tremor (shaking of fingers or a limb), slowed movement, speech changes, balance and posture problems, and rigid muscles. The combination of these symptoms often causes a characteristic slow hunched shuffling walk, illustrated in [\[link\]](#). Patients with Parkinson's disease can also exhibit psychological symptoms, such as dementia or emotional problems.

Although some patients have a form of the disease known to be caused by a single mutation, for most patients the exact causes of Parkinson's disease remain unknown: the disease likely results from a combination of genetic and environmental factors (similar to Alzheimer's disease). Post-mortem analysis of brains from Parkinson's patients shows the presence of Lewy bodies—abnormal protein clumps—in dopaminergic neurons. The prevalence of these Lewy bodies often correlates with the severity of the disease.

There is no cure for Parkinson's disease, and treatment is focused on easing symptoms. One of the most commonly prescribed drugs for Parkinson's is L-DOPA, which is a chemical that is converted into dopamine by neurons in the brain. This conversion increases the overall level of dopamine neurotransmission and can help compensate for the loss of dopaminergic neurons in the substantia nigra. Other drugs work by inhibiting the enzyme that breaks down dopamine.



Parkinson's patients often have a characteristic hunched walk.

## Neurodevelopmental Disorders

Neurodevelopmental disorders occur when the development of the nervous system is disturbed. There are several different classes of neurodevelopmental disorders. Some, like Down Syndrome, cause intellectual deficits. Others specifically affect communication, learning, or the motor system. Some disorders like autism spectrum disorder and attention deficit/hyperactivity disorder have complex symptoms.

### Autism

**Autism spectrum disorder (ASD)** is a neurodevelopmental disorder. Its severity differs from person to person. Estimates for the prevalence of the disorder have changed rapidly in the past few decades. Current estimates suggest that one in 88 children will develop the disorder. ASD is four times more prevalent in males than females.

#### Note:

Link to Learning



[This video](#) discusses possible reasons why there has been a recent increase in the number of people diagnosed with autism.

A characteristic symptom of ASD is impaired social skills. Children with autism may have difficulty making and maintaining eye contact and reading social cues. They also may have problems feeling empathy for others. Other

symptoms of ASD include repetitive motor behaviors (such as rocking back and forth), preoccupation with specific subjects, strict adherence to certain rituals, and unusual language use. Up to 30 percent of patients with ASD develop epilepsy, and patients with some forms of the disorder (like Fragile X) also have intellectual disability. Because it is a spectrum disorder, other ASD patients are very functional and have good-to-excellent language skills. Many of these patients do not feel that they suffer from a disorder and instead think that their brains just process information differently.

Except for some well-characterized, clearly genetic forms of autism (like Fragile X and Rett's Syndrome), the causes of ASD are largely unknown. Variants of several genes correlate with the presence of ASD, but for any given patient, many different mutations in different genes may be required for the disease to develop. At a general level, ASD is thought to be a disease of "incorrect" wiring. Accordingly, brains of some ASD patients lack the same level of synaptic pruning that occurs in non-affected people. In the 1990s, a research paper linked autism to a common vaccine given to children. This paper was retracted when it was discovered that the author falsified data, and follow-up studies showed no connection between vaccines and autism.

Treatment for autism usually combines behavioral therapies and interventions, along with medications to treat other disorders common to people with autism (depression, anxiety, obsessive compulsive disorder). Although early interventions can help mitigate the effects of the disease, there is currently no cure for ASD.

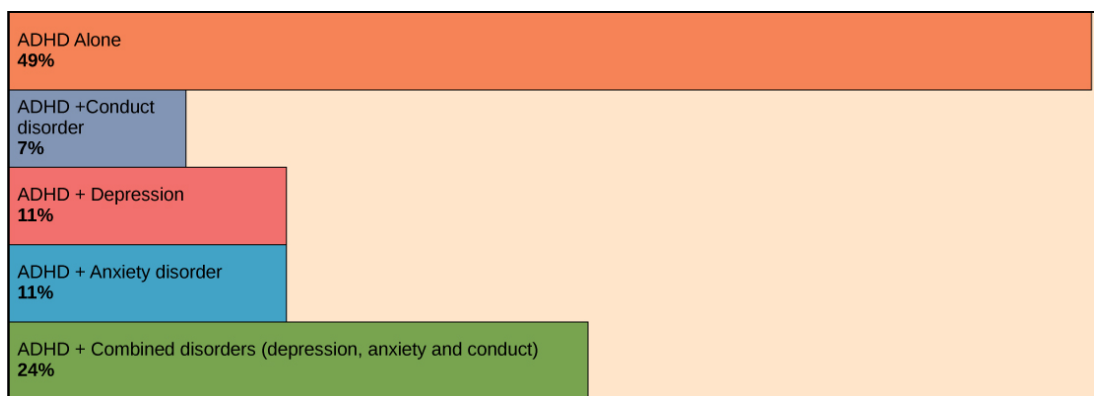
### **Attention Deficit Hyperactivity Disorder (ADHD)**

Approximately three to five percent of children and adults are affected by **attention deficit/hyperactivity disorder (ADHD)**. Like ASD, ADHD is more prevalent in males than females. Symptoms of the disorder include inattention (lack of focus), executive functioning difficulties, impulsivity, and hyperactivity beyond what is characteristic of the normal developmental stage. Some patients do not have the hyperactive component of symptoms and are diagnosed with a subtype of ADHD: attention deficit



disorder (ADD). Many people with ADHD also show comorbidity, in that they develop secondary disorders in addition to ADHD. Examples include depression or obsessive compulsive disorder (OCD). [\[link\]](#) provides some statistics concerning comorbidity with ADHD.

The cause of ADHD is unknown, although research points to a delay and dysfunction in the development of the prefrontal cortex and disturbances in neurotransmission. According to studies of twins, the disorder has a strong genetic component. There are several candidate genes that may contribute to the disorder, but no definitive links have been discovered. Environmental factors, including exposure to certain pesticides, may also contribute to the development of ADHD in some patients. Treatment for ADHD often involves behavioral therapies and the prescription of stimulant medications, which paradoxically cause a calming effect in these patients.



Many people with ADHD have one or more other neurological disorders. (credit “chart design and illustration”: modification of work by Leigh Coriale; credit “data”: Drs. Biederman and Faraone, Massachusetts General Hospital).

**Note:**  
Career Connection

## **Neurologist**

Neurologists are physicians who specialize in disorders of the nervous system. They diagnose and treat disorders such as epilepsy, stroke, dementia, nervous system injuries, Parkinson's disease, sleep disorders, and multiple sclerosis. Neurologists are medical doctors who have attended college, medical school, and completed three to four years of neurology residency.

When examining a new patient, a neurologist takes a full medical history and performs a complete physical exam. The physical exam contains specific tasks that are used to determine what areas of the brain, spinal cord, or peripheral nervous system may be damaged. For example, to check whether the hypoglossal nerve is functioning correctly, the neurologist will ask the patient to move his or her tongue in different ways. If the patient does not have full control over tongue movements, then the hypoglossal nerve may be damaged or there may be a lesion in the brainstem where the cell bodies of these neurons reside (or there could be damage to the tongue muscle itself).

Neurologists have other tools besides a physical exam they can use to diagnose particular problems in the nervous system. If the patient has had a seizure, for example, the neurologist can use electroencephalography (EEG), which involves taping electrodes to the scalp to record brain activity, to try to determine which brain regions are involved in the seizure. In suspected stroke patients, a neurologist can use a computerized tomography (CT) scan, which is a type of X-ray, to look for bleeding in the brain or a possible brain tumor. To treat patients with neurological problems, neurologists can prescribe medications or refer the patient to a neurosurgeon for surgery.

## **Note:**

Link to Learning



[This website](#) allows you to see the different tests a neurologist might use to see what regions of the nervous system may be damaged in a patient.

## Mental Illnesses

Mental illnesses are nervous system disorders that result in problems with thinking, mood, or relating with other people. These disorders are severe enough to affect a person's quality of life and often make it difficult for people to perform the routine tasks of daily living. Debilitating mental disorders plague approximately 12.5 million Americans (about 1 in 17 people) at an annual cost of more than \$300 billion. There are several types of mental disorders including schizophrenia, major depression, bipolar disorder, anxiety disorders and phobias, post-traumatic stress disorders, and obsessive-compulsive disorder (OCD), among others. The American Psychiatric Association publishes the Diagnostic and Statistical Manual of Mental Disorders (or DSM), which describes the symptoms required for a patient to be diagnosed with a particular mental disorder. Each newly released version of the DSM contains different symptoms and classifications as scientists learn more about these disorders, their causes, and how they relate to each other. A more detailed discussion of two mental illnesses—schizophrenia and major depression—is given below.

## Schizophrenia

**Schizophrenia** is a serious and often debilitating mental illness affecting one percent of people in the United States. Symptoms of the disease include the inability to differentiate between reality and imagination, inappropriate and unregulated emotional responses, difficulty thinking, and problems with

social situations. People with schizophrenia can suffer from hallucinations and hear voices; they may also suffer from delusions. Patients also have so-called “negative” symptoms like a flattened emotional state, loss of pleasure, and loss of basic drives. Many schizophrenic patients are diagnosed in their late adolescence or early 20s. The development of schizophrenia is thought to involve malfunctioning dopaminergic neurons and may also involve problems with glutamate signaling. Treatment for the disease usually requires antipsychotic medications that work by blocking dopamine receptors and decreasing dopamine neurotransmission in the brain. This decrease in dopamine can cause Parkinson’s disease-like symptoms in some patients. While some classes of antipsychotics can be quite effective at treating the disease, they are not a cure, and most patients must remain medicated for the rest of their lives.

## **Depression**

**Major depression** affects approximately 6.7 percent of the adults in the United States each year and is one of the most common mental disorders. To be diagnosed with major depressive disorder, a person must have experienced a severely depressed mood lasting longer than two weeks along with other symptoms including a loss of enjoyment in activities that were previously enjoyed, changes in appetite and sleep schedules, difficulty concentrating, feelings of worthlessness, and suicidal thoughts. The exact causes of major depression are unknown and likely include both genetic and environmental risk factors. Some research supports the “classic monoamine hypothesis,” which suggests that depression is caused by a decrease in norepinephrine and serotonin neurotransmission. One argument against this hypothesis is the fact that some antidepressant medications cause an increase in norepinephrine and serotonin release within a few hours of beginning treatment—but clinical results of these medications are not seen until weeks later. This has led to alternative hypotheses: for example, dopamine may also be decreased in depressed patients, or it may actually be an increase in norepinephrine and serotonin that causes the disease, and antidepressants force a feedback loop that decreases this release. Treatments for depression include psychotherapy, electroconvulsive therapy, deep-brain stimulation, and prescription medications. There are several classes of

antidepressant medications that work through different mechanisms. For example, monoamine oxidase inhibitors (MAO inhibitors) block the enzyme that degrades many neurotransmitters (including dopamine, serotonin, norepinephrine), resulting in increased neurotransmitter in the synaptic cleft. Selective serotonin reuptake inhibitors (SSRIs) block the reuptake of serotonin into the presynaptic neuron. This blockage results in an increase in serotonin in the synaptic cleft. Other types of drugs such as norepinephrine-dopamine reuptake inhibitors and norepinephrine-serotonin reuptake inhibitors are also used to treat depression.

## **Other Neurological Disorders**

There are several other neurological disorders that cannot be easily placed in the above categories. These include chronic pain conditions, cancers of the nervous system, epilepsy disorders, and stroke. Epilepsy and stroke are discussed below.

### **Epilepsy**

Estimates suggest that up to three percent of people in the United States will be diagnosed with **epilepsy** in their lifetime. While there are several different types of epilepsy, all are characterized by recurrent seizures. Epilepsy itself can be a symptom of a brain injury, disease, or other illness. For example, people who have intellectual disability or ASD can experience seizures, presumably because the developmental wiring malfunctions that caused their disorders also put them at risk for epilepsy. For many patients, however, the cause of their epilepsy is never identified and is likely to be a combination of genetic and environmental factors. Often, seizures can be controlled with anticonvulsant medications. However, for very severe cases, patients may undergo brain surgery to remove the brain area where seizures originate.

### **Stroke**

A stroke results when blood fails to reach a portion of the brain for a long enough time to cause damage. Without the oxygen supplied by blood flow, neurons in this brain region die. This neuronal death can cause many different symptoms—depending on the brain area affected— including headache, muscle weakness or paralysis, speech disturbances, sensory problems, memory loss, and confusion. Stroke is often caused by blood clots and can also be caused by the bursting of a weak blood vessel. Strokes are extremely common and are the third most common cause of death in the United States. On average one person experiences a stroke every 40 seconds in the United States. Approximately 75 percent of strokes occur in people older than 65. Risk factors for stroke include high blood pressure, diabetes, high cholesterol, and a family history of stroke. Smoking doubles the risk of stroke. Because a stroke is a medical emergency, patients with symptoms of a stroke should immediately go to the emergency room, where they can receive drugs that will dissolve any clot that may have formed. These drugs will not work if the stroke was caused by a burst blood vessel or if the stroke occurred more than three hours before arriving at the hospital. Treatment following a stroke can include blood pressure medication (to prevent future strokes) and (sometimes intense) physical therapy.

## **Section Summary**

Some general themes emerge from the sampling of nervous system disorders presented above. The causes for most disorders are not fully understood—at least not for all patients—and likely involve a combination of nature (genetic mutations that become risk factors) and nurture (emotional trauma, stress, hazardous chemical exposure). Because the causes have yet to be fully determined, treatment options are often lacking and only address symptoms.

## **Review Questions**

### **Exercise:**

**Problem:**

Parkinson's disease is caused by the degeneration of neurons that release \_\_\_\_\_.

- a. serotonin
- b. dopamine
- c. glutamate
- d. norepinephrine

---

**Solution:**

B

**Exercise:**

**Problem:**

\_\_\_\_\_ medications are often used to treat patients with ADHD.

- a. Tranquilizer
- b. Antibiotic
- c. Stimulant
- d. Anti-seizure

---

**Solution:**

C

**Exercise:**

**Problem:** Strokes are often caused by \_\_\_\_\_.

- a. neurodegeneration
- b. blood clots or burst blood vessels
- c. seizures
- d. viruses

---

**Solution:**

B

## Free Response

**Exercise:**

**Problem:**What are the main symptoms of Alzheimer's disease?

---

**Solution:**

Symptoms of Alzheimer's disease include disruptive memory loss, confusion about time or place, difficulties planning or executing tasks, poor judgment, and personality changes.

**Exercise:****Problem:**

What are possible treatments for patients with major depression?

---

**Solution:**

Possible treatments for patients with major depression include psychotherapy and prescription medications. MAO inhibitor drugs inhibit the breakdown of certain neurotransmitters (including dopamine, serotonin, norepinephrine) in the synaptic cleft. SSRI medications inhibit the reuptake of serotonin into the presynaptic neuron.

## Glossary

Alzheimer's disease

neurodegenerative disorder characterized by problems with memory and thinking



attention deficit hyperactivity disorder (ADHD)  
neurodevelopmental disorder characterized by difficulty maintaining attention and controlling impulses

autism spectrum disorder (ASD)  
neurodevelopmental disorder characterized by impaired social interaction and communication abilities

epilepsy  
neurological disorder characterized by recurrent seizures

major depression  
mental illness characterized by prolonged periods of sadness

neurodegenerative disorder  
nervous system disorder characterized by the progressive loss of neurological functioning, usually caused by neuron death

Parkinson's disease  
neurodegenerative disorder that affects the control of movement

schizophrenia  
mental disorder characterized by the inability to accurately perceive reality; patients often have difficulty thinking clearly and can suffer from delusions

## Introduction

class="introduction"

This shark  
uses its  
senses of  
sight,  
vibration  
(lateral-line  
system), and  
smell to  
hunt, but it  
also relies  
on its ability  
to sense the  
electric  
fields of  
prey, a sense  
not present  
in most land  
animals.

(credit:  
modificatio  
n of work  
by  
Hermanus  
Backpacker  
s Hostel,  
South  
Africa)



In more advanced animals, the senses are constantly at work, making the animal aware of stimuli—such as light, or sound, or the presence of a chemical substance in the external environment—and monitoring information about the organism's internal environment. All bilaterally symmetric animals have a sensory system, and the development of any species' sensory system has been driven by natural selection; thus, sensory systems differ among species according to the demands of their environments. The shark, unlike most fish predators, is electrosensitive—that is, sensitive to electrical fields produced by other animals in its environment. While it is helpful to this underwater predator, electrosensitivity is a sense not found in most land animals.

## Sensory Processes

By the end of this section, you will be able to:

- Identify the general and special senses in humans
- Describe three important steps in sensory perception
- Explain the concept of just-noticeable difference in sensory perception

Senses provide information about the body and its environment. Humans have five special senses: olfaction (smell), gustation (taste), equilibrium (balance and body position), vision, and hearing. Additionally, we possess general senses, also called somatosensation, which respond to stimuli like temperature, pain, pressure, and vibration. **Vestibular sensation**, which is an organism's sense of spatial orientation and balance, **proprioception** (position of bones, joints, and muscles), and the sense of limb position that is used to track **kinesthesia** (limb movement) are part of somatosensation. Although the sensory systems associated with these senses are very different, all share a common function: to convert a stimulus (such as light, or sound, or the position of the body) into an electrical signal in the nervous system. This process is called **sensory transduction**.

There are two broad types of cellular systems that perform sensory transduction. In one, a neuron works with a **sensory receptor**, a cell, or cell process that is specialized to engage with and detect a specific stimulus. Stimulation of the sensory receptor activates the associated afferent neuron, which carries information about the stimulus to the central nervous system. In the second type of sensory transduction, a sensory nerve ending responds to a stimulus in the internal or external environment: this neuron constitutes the sensory receptor. Free nerve endings can be stimulated by several different stimuli, thus showing little receptor specificity. For example, pain receptors in your gums and teeth may be stimulated by temperature changes, chemical stimulation, or pressure.

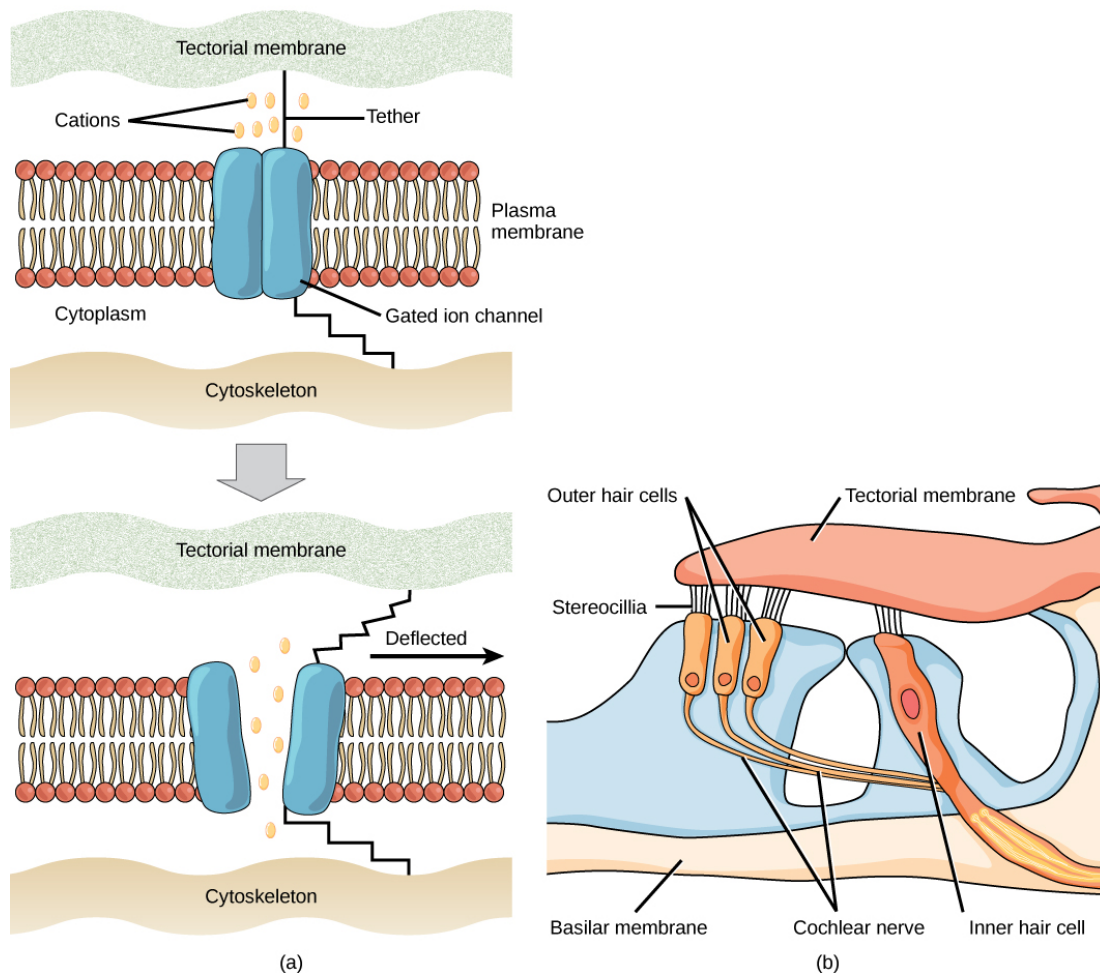
## Reception

The first step in sensation is **reception**, which is the activation of sensory receptors by stimuli such as mechanical stimuli (being bent or squished, for example), chemicals, or temperature. The receptor can then respond to the

stimuli. The region in space in which a given sensory receptor can respond to a stimulus, be it far away or in contact with the body, is that receptor's **receptive field**. Think for a moment about the differences in receptive fields for the different senses. For the sense of touch, a stimulus must come into contact with body. For the sense of hearing, a stimulus can be a moderate distance away (some baleen whale sounds can propagate for many kilometers). For vision, a stimulus can be very far away; for example, the visual system perceives light from stars at enormous distances.

## Transduction

The most fundamental function of a sensory system is the translation of a sensory signal to an electrical signal in the nervous system. This takes place at the sensory receptor, and the change in electrical potential that is produced is called the **receptor potential**. How is sensory input, such as pressure on the skin, changed to a receptor potential? In this example, a type of receptor called a **mechanoreceptor** (as shown in [\[link\]](#)) possesses specialized membranes that respond to pressure. Disturbance of these dendrites by compressing them or bending them opens gated ion channels in the plasma membrane of the sensory neuron, changing its electrical potential. Recall that in the nervous system, a positive change of a neuron's electrical potential (also called the membrane potential), depolarizes the neuron. Receptor potentials are graded potentials: the magnitude of these graded (receptor) potentials varies with the strength of the stimulus. If the magnitude of depolarization is sufficient (that is, if membrane potential reaches a threshold), the neuron will fire an action potential. In most cases, the correct stimulus impinging on a sensory receptor will drive membrane potential in a positive direction, although for some receptors, such as those in the visual system, this is not always the case.



(a) Mechanosensitive ion channels are gated ion channels that respond to mechanical deformation of the plasma membrane. A mechanosensitive channel is connected to the plasma membrane and the cytoskeleton by hair-like tethers. When pressure causes the extracellular matrix to move, the channel opens, allowing ions to enter or exit the cell. (b) Stereocilia in the human ear are connected to mechanosensitive ion channels.

When a sound causes the stereocilia to move, mechanosensitive ion channels transduce the signal to the cochlear nerve.

Sensory receptors for different senses are very different from each other, and they are specialized according to the type of stimulus they sense: they

have receptor specificity. For example, touch receptors, light receptors, and sound receptors are each activated by different stimuli. Touch receptors are not sensitive to light or sound; they are sensitive only to touch or pressure. However, stimuli may be combined at higher levels in the brain, as happens with olfaction, contributing to our sense of taste.

## **Encoding and Transmission of Sensory Information**

Four aspects of sensory information are encoded by sensory systems: the type of stimulus, the location of the stimulus in the receptive field, the duration of the stimulus, and the relative intensity of the stimulus. Thus, action potentials transmitted over a sensory receptor's afferent axons encode one type of stimulus, and this segregation of the senses is preserved in other sensory circuits. For example, auditory receptors transmit signals over their own dedicated system, and electrical activity in the axons of the auditory receptors will be interpreted by the brain as an auditory stimulus—a sound.

The intensity of a stimulus is often encoded in the rate of action potentials produced by the sensory receptor. Thus, an intense stimulus will produce a more rapid train of action potentials, and reducing the stimulus will likewise slow the rate of production of action potentials. A second way in which intensity is encoded is by the number of receptors activated. An intense stimulus might initiate action potentials in a large number of adjacent receptors, while a less intense stimulus might stimulate fewer receptors. Integration of sensory information begins as soon as the information is received in the CNS, and the brain will further process incoming signals.

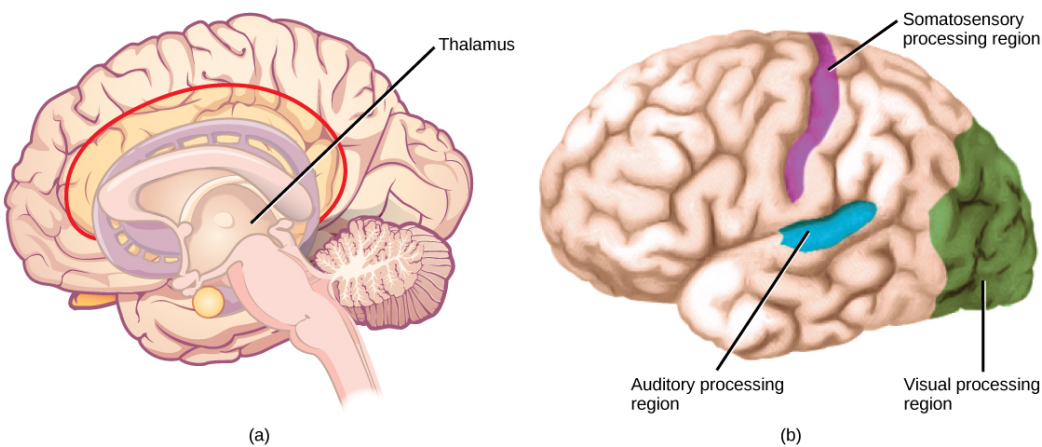
## **Perception**

**Perception** is an individual's interpretation of a sensation. Although perception relies on the activation of sensory receptors, perception happens not at the level of the sensory receptor, but at higher levels in the nervous system, in the brain. The brain distinguishes sensory stimuli through a

sensory pathway: action potentials from sensory receptors travel along neurons that are dedicated to a particular stimulus. These neurons are dedicated to that particular stimulus and synapse with particular neurons in the brain or spinal cord.

All sensory signals, except those from the olfactory system, are transmitted through the central nervous system and are routed to the thalamus and to the appropriate region of the cortex. Recall that the thalamus is a structure in the forebrain that serves as a clearinghouse and relay station for sensory (as well as motor) signals. When the sensory signal exits the thalamus, it is conducted to the specific area of the cortex ([link](#)) dedicated to processing that particular sense.

How are neural signals interpreted? Interpretation of sensory signals between individuals of the same species is largely similar, owing to the inherited similarity of their nervous systems; however, there are some individual differences. A good example of this is individual tolerances to a painful stimulus, such as dental pain, which certainly differ.



In humans, with the exception of olfaction, all sensory signals are routed from the (a) thalamus to (b) final processing regions in the cortex of the brain. (credit b: modification of work by Polina Tishina)



**Note:****Scientific Method Connection****Just-Noticeable Difference**

It is easy to differentiate between a one-pound bag of rice and a two-pound bag of rice. There is a one-pound difference, and one bag is twice as heavy as the other. However, would it be as easy to differentiate between a 20- and a 21-pound bag?

**Question:** What is the smallest detectible weight difference between a one-pound bag of rice and a larger bag? What is the smallest detectible difference between a 20-pound bag and a larger bag? In both cases, at what weights are the differences detected? This smallest detectible difference in stimuli is known as the just-noticeable difference (JND).

**Background:** Research background literature on JND and on Weber's Law, a description of a proposed mathematical relationship between the overall magnitude of the stimulus and the JND. You will be testing JND of different weights of rice in bags. Choose a convenient increment that is to be stepped through while testing. For example, you could choose 10 percent increments between one and two pounds (1.1, 1.2, 1.3, 1.4, and so on) or 20 percent increments (1.2, 1.4, 1.6, and 1.8).

**Hypothesis:** Develop a hypothesis about JND in terms of percentage of the whole weight being tested (such as "the JND between the two small bags and between the two large bags is proportionally the same," or "... is not proportionally the same.") So, for the first hypothesis, if the JND between the one-pound bag and a larger bag is 0.2 pounds (that is, 20 percent; 1.0 pound feels the same as 1.1 pounds, but 1.0 pound feels less than 1.2 pounds), then the JND between the 20-pound bag and a larger bag will also be 20 percent. (So, 20 pounds feels the same as 22 pounds or 23 pounds, but 20 pounds feels less than 24 pounds.)

**Test the hypothesis:** Enlist 24 participants, and split them into two groups of 12. To set up the demonstration, assuming a 10 percent increment was selected, have the first group be the one-pound group. As a counter-balancing measure against a systematic error, however, six of the first group will compare one pound to two pounds, and step down in weight (1.0 to 2.0, 1.0 to 1.9, and so on.), while the other six will step up (1.0 to 1.1, 1.0 to 1.2, and so on). Apply the same principle to the 20-pound group (20 to 40, 20 to 38, and so on, and 20 to 22, 20 to 24, and so on). Given the large difference between 20 and 40 pounds, you may wish to use 30

pounds as your larger weight. In any case, use two weights that are easily detectable as different.

**Record the observations:** Record the data in a table similar to the table below. For the one-pound and 20-pound groups (base weights) record a plus sign (+) for each participant that detects a difference between the base weight and the step weight. Record a minus sign (-) for each participant that finds no difference. If one-tenth steps were not used, then replace the steps in the “Step Weight” columns with the step you are using.

**Results of JND Testing (+ = difference; – = no difference)**

Step Weight	One pound	20 pounds	Step Weight
1.1			22
1.2			24
1.3			26
1.4			28
1.5			30
1.6			32
1.7			34
1.8			36
1.9			38
2.0			40

**Analyze the data/report the results:** What step weight did all participants find to be equal with one-pound base weight? What about the 20-pound group?

**Draw a conclusion:** Did the data support the hypothesis? Are the final weights proportionally the same? If not, why not? Do the findings adhere to Weber's Law? Weber's Law states that the concept that a just-noticeable difference in a stimulus is proportional to the magnitude of the original stimulus.

## Section Summary

A sensory activation occurs when a physical or chemical stimulus is processed into a neural signal (sensory transduction) by a sensory receptor. Perception is an individual interpretation of a sensation and is a brain function. Humans have special senses: olfaction, gustation, equilibrium, and hearing, plus the general senses of somatosensation.

Sensory receptors are either specialized cells associated with sensory neurons or the specialized ends of sensory neurons that are a part of the peripheral nervous system, and they are used to receive information about the environment (internal or external). Each sensory receptor is modified for the type of stimulus it detects. For example, neither gustatory receptors nor auditory receptors are sensitive to light. Each sensory receptor is responsive to stimuli within a specific region in space, which is known as that receptor's receptive field. The most fundamental function of a sensory system is the translation of a sensory signal to an electrical signal in the nervous system.

All sensory signals, except those from the olfactory system, enter the central nervous system and are routed to the thalamus. When the sensory signal exits the thalamus, it is conducted to the specific area of the cortex dedicated to processing that particular sense.

## Review Questions

**Exercise:**

**Problem:** Where does perception occur?

- a. spinal cord
- b. cerebral cortex
- c. receptors
- d. thalamus

---

**Solution:**

B

**Exercise:**

**Problem:**

If a person's cold receptors no longer convert cold stimuli into sensory signals, that person has a problem with the process of \_\_\_\_\_.

- a. reception
- b. transmission
- c. perception
- d. transduction

---

**Solution:**

D

**Exercise:**

**Problem:**

After somatosensory transduction, the sensory signal travels through the brain as a(n) \_\_\_\_\_ signal.

- a. electrical
- b. pressure

- c. optical
  - d. thermal
- 

**Solution:**

A

## Free Response

**Exercise:**

**Problem:**

If a person sustains damage to axons leading from sensory receptors to the central nervous system, which step or steps of sensory perception will be affected?

---

**Solution:**

Transmission of sensory information from the receptor to the central nervous system will be impaired, and thus, perception of stimuli, which occurs in the brain, will be halted.

**Exercise:**

**Problem:**

In what way does the overall magnitude of a stimulus affect the just-noticeable difference in the perception of that stimulus?

---

**Solution:**

The just-noticeable difference is a fraction of the overall magnitude of the stimulus and seems to be a relatively fixed proportion (such as 10 percent) whether the stimulus is large (such as a very heavy object) or small (such as a very light object).

## **Glossary**

kinesthesia

sense of body movement

mechanoreceptor

sensory receptor modified to respond to mechanical disturbance such as being bent, touch, pressure, motion, and sound

perception

individual interpretation of a sensation; a brain function

proprioception

sense of limb position; used to track kinesthesia

reception

receipt of a signal (such as light or sound) by sensory receptors

receptive field

region in space in which a stimulus can activate a given sensory receptor

receptor potential

membrane potential in a sensory receptor in response to detection of a stimulus

sensory receptor

specialized neuron or other cells associated with a neuron that is modified to receive specific sensory input

sensory transduction

conversion of a sensory stimulus into electrical energy in the nervous system by a change in the membrane potential

vestibular sense

sense of spatial orientation and balance

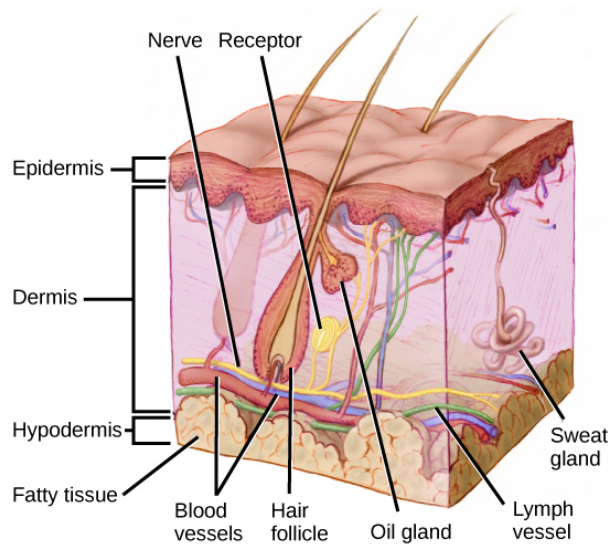
## Somatosensation

By the end of this section, you will be able to:

- Describe four important mechanoreceptors in human skin
- Describe the topographical distribution of somatosensory receptors between glabrous and hairy skin
- Explain why the perception of pain is subjective

Somatosensation is a mixed sensory category and includes all sensation received from the skin and mucous membranes, as well from the limbs and joints. Somatosensation is also known as tactile sense, or more familiarly, as the sense of touch. Somatosensation occurs all over the exterior of the body and at some interior locations as well. A variety of receptor types—embedded in the skin, mucous membranes, muscles, joints, internal organs, and cardiovascular system—play a role.

Recall that the epidermis is the outermost layer of skin in mammals. It is relatively thin, is composed of keratin-filled cells, and has no blood supply. The epidermis serves as a barrier to water and to invasion by pathogens. Below this, the much thicker dermis contains blood vessels, sweat glands, hair follicles, lymph vessels, and lipid-secreting sebaceous glands ([\[link\]](#)). Below the epidermis and dermis is the subcutaneous tissue, or hypodermis, the fatty layer that contains blood vessels, connective tissue, and the axons of sensory neurons. The hypodermis, which holds about 50 percent of the body's fat, attaches the dermis to the bone and muscle, and supplies nerves and blood vessels to the dermis.



Mammalian skin has three layers: an epidermis, a dermis, and a hypodermis. (credit: modification of work by Don Bliss, National Cancer Institute)

## Somatosensory Receptors

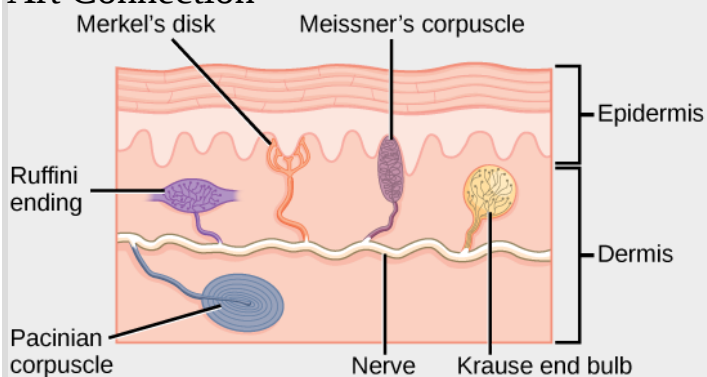
Sensory receptors are classified into five categories: mechanoreceptors, thermoreceptors, proprioceptors, pain receptors, and chemoreceptors. These categories are based on the nature of stimuli each receptor class transduces. What is commonly referred to as “touch” involves more than one kind of stimulus and more than one kind of receptor. Mechanoreceptors in the skin are described as encapsulated (that is, surrounded by a capsule) or unencapsulated (a group that includes free nerve endings). A **free nerve ending**, as its name implies, is an unencapsulated dendrite of a sensory neuron. Free nerve endings are the most common nerve endings in skin, and they extend into the middle of the epidermis. Free nerve endings are sensitive to painful stimuli, to hot and cold, and to light touch. They are slow to adjust to a stimulus and so are less sensitive to abrupt changes in stimulation.



There are three classes of mechanoreceptors: tactile, proprioceptors, and baroreceptors. Mechanoreceptors sense stimuli due to physical deformation of their plasma membranes. They contain mechanically gated ion channels whose gates open or close in response to pressure, touch, stretching, and sound.” There are four primary tactile mechanoreceptors in human skin: Merkel’s disks, Meissner’s corpuscles, Ruffini endings, and Pacinian corpuscle; two are located toward the surface of the skin and two are located deeper. A fifth type of mechanoreceptor, Krause end bulbs, are found only in specialized regions. **Merkel’s disks** (shown in [\[link\]](#)) are found in the upper layers of skin near the base of the epidermis, both in skin that has hair and on **glabrous** skin, that is, the hairless skin found on the palms and fingers, the soles of the feet, and the lips of humans and other primates. Merkel’s disks are densely distributed in the fingertips and lips. They are slow-adapting, encapsulated nerve endings, and they respond to light touch. Light touch, also known as discriminative touch, is a light pressure that allows the location of a stimulus to be pinpointed. The receptive fields of Merkel’s disks are small with well-defined borders. That makes them finely sensitive to edges and they come into use in tasks such as typing on a keyboard.

### Note:

#### Art Connection



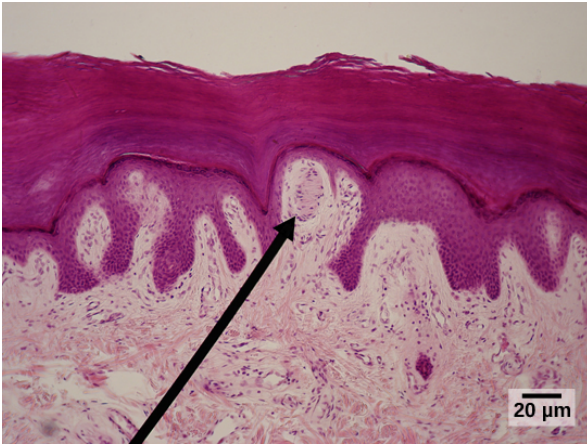
Four of the primary mechanoreceptors in human skin are shown. Merkel’s disks, which are unencapsulated, respond to light touch. Meissner’s

corpuscles, Ruffini endings, Pacinian corpuscles, and Krause end bulbs are all encapsulated. Meissner's corpuscles respond to touch and low-frequency vibration. Ruffini endings detect stretch, deformation within joints, and warmth. Pacinian corpuscles detect transient pressure and high-frequency vibration. Krause end bulbs detect cold.

Which of the following statements about mechanoreceptors is false?

- a. Pacini corpuscles are found in both glabrous and hairy skin.
- b. Merkel's disks are abundant on the fingertips and lips.
- c. Ruffini endings are encapsulated mechanoreceptors.
- d. Meissner's corpuscles extend into the lower dermis.

**Meissner's corpuscles**, (shown in [\[link\]](#)) also known as tactile corpuscles, are found in the upper dermis, but they project into the epidermis. They, too, are found primarily in the glabrous skin on the fingertips and eyelids. They respond to fine touch and pressure, but they also respond to low-frequency vibration or flutter. They are rapidly adapting, fluid-filled, encapsulated neurons with small, well-defined borders and are responsive to fine details. Like Merkel's disks, Meissner's corpuscles are not as plentiful in the palms as they are in the fingertips.

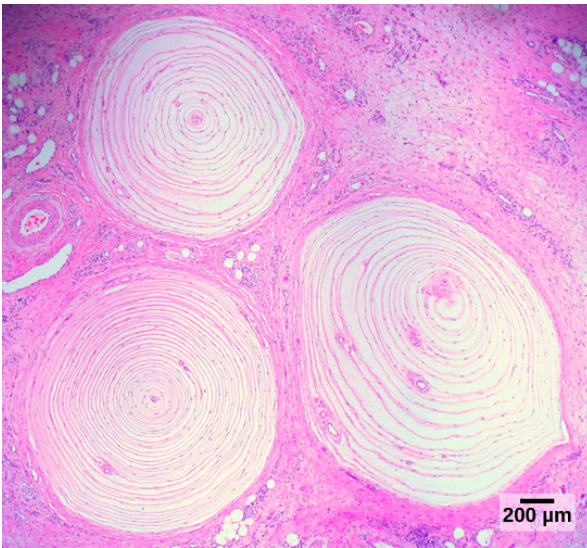


Meissner corpuscles in the fingertips, such as the one viewed here using bright field light microscopy, allow for touch discrimination of fine detail. (credit: modification of work by "Wbensmith"/Wikimedia Commons; scale-bar data from Matt Russell)

Deeper in the epidermis, near the base, are **Ruffini endings**, which are also known as bulbous corpuscles. They are found in both glabrous and hairy skin. These are slow-adapting, encapsulated mechanoreceptors that detect skin stretch and deformations within joints, so they provide valuable feedback for gripping objects and controlling finger position and movement. Thus, they also contribute to proprioception and kinesthesia. Ruffini endings also detect warmth. Note that these warmth detectors are situated deeper in the skin than are the cold detectors. It is not surprising, then, that humans detect cold stimuli before they detect warm stimuli.

**Pacinian corpuscles** (seen in [\[link\]](#)) are located deep in the dermis of both glabrous and hairy skin and are structurally similar to Meissner's corpuscles; they are found in the bone periosteum, joint capsules, pancreas

and other viscera, breast, and genitals. They are rapidly adapting mechanoreceptors that sense deep transient (but not prolonged) pressure and high-frequency vibration. Pacinian receptors detect pressure and vibration by being compressed, stimulating their internal dendrites. There are fewer Pacinian corpuscles and Ruffini endings in skin than there are Merkel's disks and Meissner's corpuscles.



Pacinian corpuscles, such as these visualized using bright field light microscopy, detect pressure (touch) and high-frequency vibration. (credit: modification of work by Ed Uthman; scale-bar data from Matt Russell)

In proprioception, proprioceptive and kinesthetic signals travel through myelinated afferent neurons running from the spinal cord to the medulla. Neurons are not physically connected, but communicate via neurotransmitters secreted into synapses or “gaps” between communicating

neurons. Once in the medulla, the neurons continue carrying the signals to the thalamus.

**Muscle spindles** are stretch receptors that detect the amount of stretch, or lengthening of muscles. Related to these are **Golgi tendon organs**, which are tension receptors that detect the force of muscle contraction.

Proprioceptive and kinesthetic signals come from limbs. Unconscious proprioceptive signals run from the spinal cord to the cerebellum, the brain region that coordinates muscle contraction, rather than to the thalamus, like most other sensory information.

Baroreceptors detect pressure changes in an organ. They are found in the walls of the carotid artery and the aorta where they monitor blood pressure, and in the lungs where they detect the degree of lung expansion. Stretch receptors are found at various sites in the digestive and urinary systems.

In addition to these two types of deeper receptors, there are also rapidly adapting hair receptors, which are found on nerve endings that wrap around the base of hair follicles. There are a few types of hair receptors that detect slow and rapid hair movement, and they differ in their sensitivity to movement. Some hair receptors also detect skin deflection, and certain rapidly adapting hair receptors allow detection of stimuli that have not yet touched the skin.

## **Integration of Signals from Mechanoreceptors**

The configuration of the different types of receptors working in concert in human skin results in a very refined sense of touch. The nociceptive receptors—those that detect pain—are located near the surface. Small, finely calibrated mechanoreceptors—Merkel's disks and Meissner's corpuscles—are located in the upper layers and can precisely localize even gentle touch. The large mechanoreceptors—Pacinian corpuscles and Ruffini endings—are located in the lower layers and respond to deeper touch. (Consider that the deep pressure that reaches those deeper receptors would not need to be finely localized.) Both the upper and lower layers of the skin hold rapidly and slowly adapting receptors. Both primary somatosensory

cortex and secondary cortical areas are responsible for processing the complex picture of stimuli transmitted from the interplay of mechanoreceptors.

## **Density of Mechanoreceptors**

The distribution of touch receptors in human skin is not consistent over the body. In humans, touch receptors are less dense in skin covered with any type of hair, such as the arms, legs, torso, and face. Touch receptors are denser in glabrous skin (the type found on human fingertips and lips, for example), which is typically more sensitive and is thicker than hairy skin (4 to 5 mm versus 2 to 3 mm).

How is receptor density estimated in a human subject? The relative density of pressure receptors in different locations on the body can be demonstrated experimentally using a two-point discrimination test. In this demonstration, two sharp points, such as two thumbtacks, are brought into contact with the subject's skin (though not hard enough to cause pain or break the skin). The subject reports if he or she feels one point or two points. If the two points are felt as one point, it can be inferred that the two points are both in the receptive field of a single sensory receptor. If two points are felt as two separate points, each is in the receptive field of two separate sensory receptors. The points could then be moved closer and re-tested until the subject reports feeling only one point, and the size of the receptive field of a single receptor could be estimated from that distance.

## **Thermoreception**

In addition to Krause end bulbs that detect cold and Ruffini endings that detect warmth, there are different types of cold receptors on some free nerve endings: thermoreceptors, located in the dermis, skeletal muscles, liver, and hypothalamus, that are activated by different temperatures. Their pathways into the brain run from the spinal cord through the thalamus to the primary somatosensory cortex. Warmth and cold information from the face travels through one of the cranial nerves to the brain. You know from

experience that a tolerably cold or hot stimulus can quickly progress to a much more intense stimulus that is no longer tolerable. Any stimulus that is too intense can be perceived as pain because temperature sensations are conducted along the same pathways that carry pain sensations

## **Pain**

Pain is the name given to **nociception**, which is the neural processing of injurious stimuli in response to tissue damage. Pain is caused by true sources of injury, such as contact with a heat source that causes a thermal burn or contact with a corrosive chemical. But pain also can be caused by harmless stimuli that mimic the action of damaging stimuli, such as contact with capsaicins, the compounds that cause peppers to taste hot and which are used in self-defense pepper sprays and certain topical medications. Peppers taste “hot” because the protein receptors that bind capsaicin open the same calcium channels that are activated by warm receptors.

Nociception starts at the sensory receptors, but pain, inasmuch as it is the perception of nociception, does not start until it is communicated to the brain. There are several nociceptive pathways to and through the brain. Most axons carrying nociceptive information into the brain from the spinal cord project to the thalamus (as do other sensory neurons) and the neural signal undergoes final processing in the primary somatosensory cortex. Interestingly, one nociceptive pathway projects not to the thalamus but directly to the hypothalamus in the forebrain, which modulates the cardiovascular and neuroendocrine functions of the autonomic nervous system. Recall that threatening—or painful—stimuli stimulate the sympathetic branch of the visceral sensory system, readying a fight-or-flight response.

### **Note:**

[Link to Learning](#)



View this [video](https://www.openstaxcollege.org/l/nociceptive) that animates the five phases of nociceptive pain.  
<https://www.openstaxcollege.org/l/nociceptive>

## Section Summary

Somatosensation includes all sensation received from the skin and mucous membranes, as well as from the limbs and joints. Somatosensation occurs all over the exterior of the body and at some interior locations as well, and a variety of receptor types, embedded in the skin and mucous membranes, play a role.

There are several types of specialized sensory receptors. Rapidly adapting free nerve endings detect nociception, hot and cold, and light touch. Slowly adapting, encapsulated Merkel's disks are found in fingertips and lips, and respond to light touch. Meissner's corpuscles, found in glabrous skin, are rapidly adapting, encapsulated receptors that detect touch, low-frequency vibration, and flutter. Ruffini endings are slowly adapting, encapsulated receptors that detect skin stretch, joint activity, and warmth. Hair receptors are rapidly adapting nerve endings wrapped around the base of hair follicles that detect hair movement and skin deflection. Finally, Pacinian corpuscles are encapsulated, rapidly adapting receptors that detect transient pressure and high-frequency vibration.

## Art Connections

### Exercise:



**Problem:**

[\[link\]](#) Which of the following statements about mechanoreceptors is false?

- a. Pacini corpuscles are found in both glabrous and hairy skin.
- b. Merkel's disks are abundant on the fingertips and lips.
- c. Ruffini endings are encapsulated mechanoreceptors.
- d. Meissner's corpuscles extend into the lower dermis.

---

**Solution:**

[\[link\]](#) D

**Review Questions****Exercise:****Problem:**

\_\_\_\_\_ are found only in \_\_\_\_\_ skin, and detect skin deflection.

- a. Meissner's corpuscles: hairy
- b. Merkel's disks: glabrous
- c. hair receptors: hairy
- d. Krause end bulbs: hairy

---

**Solution:**

B

**Exercise:****Problem:**

If you were to burn your epidermis, what receptor type would you most likely burn?

- a. free nerve endings
  - b. Ruffini endings
  - c. Pacinian corpuscle
  - d. hair receptors
- 

**Solution:**

A

## Free Response

**Exercise:**

**Problem:**

What can be inferred about the relative sizes of the areas of cortex that process signals from skin not densely innervated with sensory receptors and skin that is densely innervated with sensory receptors?

---

**Solution:**

The cortical areas serving skin that is densely innervated likely are larger than those serving skin that is less densely innervated.

## Glossary

free nerve ending

ending of an afferent neuron that lacks a specialized structure for detection of sensory stimuli; some respond to touch, pain, or temperature

glabrous

describes the non-hairy skin found on palms and fingers, soles of feet, and lips of humans and other primates

Golgi tendon organ

muscular proprioceptive tension receptor that provides the sensory component of the Golgi tendon reflex

Meissner's corpuscle

(also, tactile corpuscle) encapsulated, rapidly-adapting mechanoreceptor in the skin that responds to light touch

Merkel's disc

unencapsulated, slowly-adapting mechanoreceptor in the skin that responds to touch

muscle spindle

proprioceptive stretch receptor that lies within a muscle and that shortens the muscle to an optimal length for efficient contraction

nociception

neural processing of noxious (such as damaging) stimuli

Pacinian corpuscle

encapsulated mechanoreceptor in the skin that responds to deep pressure and vibration

Ruffini ending

(also, bulbous corpuscle) slowly-adapting mechanoreceptor in the skin that responds to skin stretch and joint position

## Taste and Smell

By the end of this section, you will be able to:

- Explain in what way smell and taste stimuli differ from other sensory stimuli
- Identify the five primary tastes that can be distinguished by humans
- Explain in anatomical terms why a dog's sense of smell is more acute than a human's

Taste, also called **gustation**, and smell, also called **olfaction**, are the most interconnected senses in that both involve molecules of the stimulus entering the body and bonding to receptors. Smell lets an animal sense the presence of food or other animals—whether potential mates, predators, or prey—or other chemicals in the environment that can impact their survival. Similarly, the sense of taste allows animals to discriminate between types of foods. While the value of a sense of smell is obvious, what is the value of a sense of taste? Different tasting foods have different attributes, both helpful and harmful. For example, sweet-tasting substances tend to be highly caloric, which could be necessary for survival in lean times. Bitterness is associated with toxicity, and sourness is associated with spoiled food. Salty foods are valuable in maintaining homeostasis by helping the body retain water and by providing ions necessary for cells to function.

## Tastes and Odors

Both taste and odor stimuli are molecules taken in from the environment. The primary tastes detected by humans are sweet, sour, bitter, salty and umami. The first four tastes need little explanation. The identification of **umami** as a fundamental taste occurred fairly recently—it was identified in 1908 by Japanese scientist Kikunae Ikeda while he worked with seaweed broth, but it was not widely accepted as a taste that could be physiologically distinguished until many years later. The taste of umami, also known as savoriness, is attributable to the taste of the amino acid L-glutamate. In fact, monosodium glutamate, or MSG, is often used in cooking to enhance the savory taste of certain foods. What is the adaptive value of being able to distinguish umami? Savory substances tend to be high in protein.

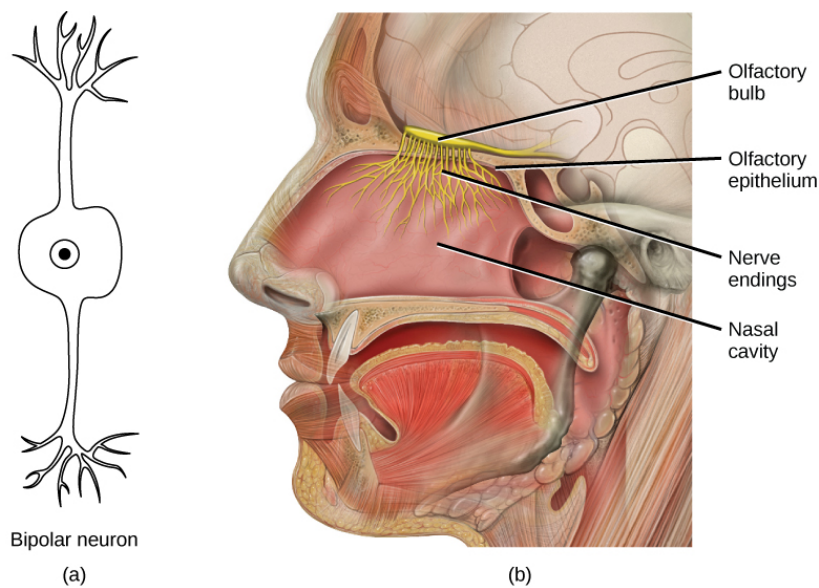
All odors that we perceive are molecules in the air we breathe. If a substance does not release molecules into the air from its surface, it has no smell. And if a human or other animal does not have a receptor that recognizes a specific molecule, then that molecule has no smell. Humans have about 350 olfactory receptor subtypes that work in various combinations to allow us to sense about 10,000 different odors. Compare that to mice, for example, which have about 1,300 olfactory receptor types, and therefore probably sense more odors. Both odors and tastes involve molecules that stimulate specific chemoreceptors. Although humans commonly distinguish taste as one sense and smell as another, they work together to create the perception of flavor. A person's perception of flavor is reduced if he or she has congested nasal passages.

## Reception and Transduction

**Odorants** (odor molecules) enter the nose and dissolve in the olfactory epithelium, the mucosa at the back of the nasal cavity (as illustrated in [\[link\]](#)). The **olfactory epithelium** is a collection of specialized olfactory receptors in the back of the nasal cavity that spans an area about 5 cm<sup>2</sup> in humans. Recall that sensory cells are neurons. An **olfactory receptor**, which is a dendrite of a specialized neuron, responds when it binds certain molecules inhaled from the environment by sending impulses directly to the olfactory bulb of the brain. Humans have about 12 million olfactory receptors, distributed among hundreds of different receptor types that respond to different odors. Twelve million seems like a large number of receptors, but compare that to other animals: rabbits have about 100 million, most dogs have about 1 billion, and bloodhounds—dogs selectively bred for their sense of smell—have about 4 billion. The overall size of the olfactory epithelium also differs between species, with that of bloodhounds, for example, being many times larger than that of humans.

Olfactory neurons are **bipolar neurons** (neurons with two processes from the cell body). Each neuron has a single dendrite buried in the olfactory epithelium, and extending from this dendrite are 5 to 20 receptor-laden, hair-like cilia that trap odorant molecules. The sensory receptors on the cilia are proteins, and it is the variations in their amino acid chains that make the receptors sensitive to different odorants. Each olfactory sensory neuron has

only one type of receptor on its cilia, and the receptors are specialized to detect specific odorants, so the bipolar neurons themselves are specialized. When an odorant binds with a receptor that recognizes it, the sensory neuron associated with the receptor is stimulated. Olfactory stimulation is the only sensory information that directly reaches the cerebral cortex, whereas other sensations are relayed through the thalamus.



In the human olfactory system, (a) bipolar olfactory neurons extend from (b) the olfactory epithelium, where olfactory receptors are located, to the olfactory bulb. (credit: modification of work by Patrick J. Lynch, medical illustrator; C. Carl Jaffe, MD, cardiologist)

### **Note:**

Evolution Connection

**Pheromones**

A **pheromone** is a chemical released by an animal that affects the behavior or physiology of animals of the same species. Pheromonal signals can have profound effects on animals that inhale them, but pheromones apparently are not consciously perceived in the same way as other odors. There are several different types of pheromones, which are released in urine or as glandular secretions. Certain pheromones are attractants to potential mates, others are repellants to potential competitors of the same sex, and still others play roles in mother-infant attachment. Some pheromones can also influence the timing of puberty, modify reproductive cycles, and even prevent embryonic implantation. While the roles of pheromones in many nonhuman species are important, pheromones have become less important in human behavior over evolutionary time compared to their importance to organisms with more limited behavioral repertoires.

The vomeronasal organ (VNO, or Jacobson's organ) is a tubular, fluid-filled, olfactory organ present in many vertebrate animals that sits adjacent to the nasal cavity. It is very sensitive to pheromones and is connected to the nasal cavity by a duct. When molecules dissolve in the mucosa of the nasal cavity, they then enter the VNO where the pheromone molecules among them bind with specialized pheromone receptors. Upon exposure to pheromones from their own species or others, many animals, including cats, may display the flehmen response (shown in [\[link\]](#)), a curling of the upper lip that helps pheromone molecules enter the VNO.

Pheromonal signals are sent, not to the main olfactory bulb, but to a different neural structure that projects directly to the amygdala (recall that the amygdala is a brain center important in emotional reactions, such as fear). The pheromonal signal then continues to areas of the hypothalamus that are key to reproductive physiology and behavior. While some scientists assert that the VNO is apparently functionally vestigial in humans, even though there is a similar structure located near human nasal cavities, others are researching it as a possible functional system that may, for example, contribute to synchronization of menstrual cycles in women living in close proximity.

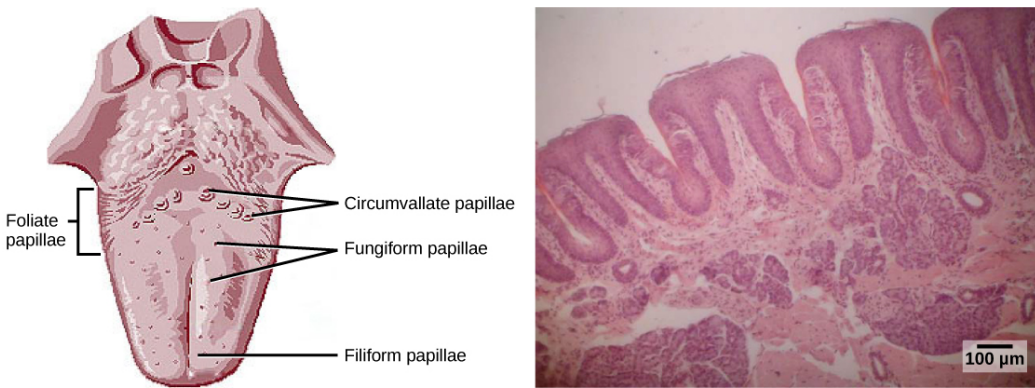


The flehmen response in this tiger results in the curling of the upper lip and helps airborne pheromone molecules enter the vomeronasal organ. (credit: modification of work by "chadh"/Flickr)

## Taste

Detecting a taste (gustation) is fairly similar to detecting an odor (olfaction), given that both taste and smell rely on chemical receptors being stimulated by certain molecules. The primary organ of taste is the taste bud. A **taste bud** is a cluster of gustatory receptors (taste cells) that are located within the bumps on the tongue called **papillae** (singular: papilla) (illustrated in [\[link\]](#)). There are several structurally distinct papillae. Filiform papillae, which are located across the tongue, are tactile, providing friction that helps the tongue move substances, and contain no taste cells. In contrast, fungiform papillae, which are located mainly on the anterior two-thirds of the tongue, each contain one to eight taste buds and also have receptors for pressure and temperature. The large circumvallate papillae contain up to 100 taste buds and form a V near the posterior margin of the tongue.

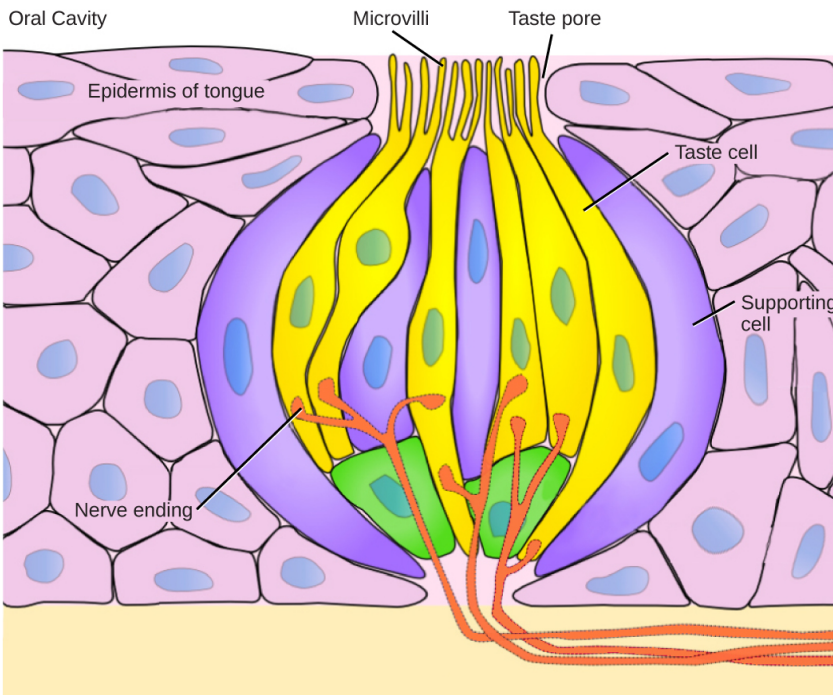




(a) Foliate, circumvallate, and fungiform papillae are located on different regions of the tongue. (b) Foliate papillae are prominent protrusions on this light micrograph. (credit a: modification of work by NCI; scale-bar data from Matt Russell)

In addition to those two types of chemically and mechanically sensitive papillae are foliate papillae—leaf-like papillae located in parallel folds along the edges and toward the back of the tongue, as seen in the [\[link\]](#) micrograph. Foliate papillae contain about 1,300 taste buds within their folds. Finally, there are circumvallate papillae, which are wall-like papillae in the shape of an inverted “V” at the back of the tongue. Each of these papillae is surrounded by a groove and contains about 250 taste buds.

Each taste bud’s taste cells are replaced every 10 to 14 days. These are elongated cells with hair-like processes called microvilli at the tips that extend into the taste bud pore (illustrate in [\[link\]](#)). Food molecules (**tastants**) are dissolved in saliva, and they bind with and stimulate the receptors on the microvilli. The receptors for tastants are located across the outer portion and front of the tongue, outside of the middle area where the filiform papillae are most prominent.



Pores in the tongue allow tastants to enter taste pores in the tongue. (credit: modification of work by Vincenzo Rizzo)

In humans, there are five primary tastes, and each taste has only one corresponding type of receptor. Thus, like olfaction, each receptor is specific to its stimulus (tastant). Transduction of the five tastes happens through different mechanisms that reflect the molecular composition of the tastant. A salty tastant (containing NaCl) provides the sodium ions ( $\text{Na}^+$ ) that enter the taste neurons and excite them directly. Sour tastants are acids and belong to the thermoreceptor protein family. Binding of an acid or other sour-tasting molecule triggers a change in the ion channel and these increase hydrogen ion ( $\text{H}^+$ ) concentrations in the taste neurons, thus depolarizing them. Sweet, bitter, and umami tastants require a G-protein coupled receptor. These tastants bind to their respective receptors, thereby exciting the specialized neurons associated with them.

Both tasting abilities and sense of smell change with age. In humans, the senses decline dramatically by age 50 and continue to decline. A child may

find a food to be too spicy, whereas an elderly person may find the same food to be bland and unappetizing.

**Note:**

Link to Learning



View this [animation](#) that shows how the sense of taste works.

## Smell and Taste in the Brain

Olfactory neurons project from the olfactory epithelium to the olfactory bulb as thin, unmyelinated axons. The **olfactory bulb** is composed of neural clusters called **glomeruli**, and each glomerulus receives signals from one type of olfactory receptor, so each glomerulus is specific to one odorant. From glomeruli, olfactory signals travel directly to the olfactory cortex and then to the frontal cortex and the thalamus. Recall that this is a different path from most other sensory information, which is sent directly to the thalamus before ending up in the cortex. Olfactory signals also travel directly to the amygdala, thereafter reaching the hypothalamus, thalamus, and frontal cortex. The last structure that olfactory signals directly travel to is a cortical center in the temporal lobe structure important in spatial, autobiographical, declarative, and episodic memories. Olfaction is finally processed by areas of the brain that deal with memory, emotions, reproduction, and thought.

Taste neurons project from taste cells in the tongue, esophagus, and palate to the medulla, in the brainstem. From the medulla, taste signals travel to the thalamus and then to the primary gustatory cortex. Information from

different regions of the tongue is segregated in the medulla, thalamus, and cortex.

## Section Summary

There are five primary tastes in humans: sweet, sour, bitter, salty, and umami. Each taste has its own receptor type that responds only to that taste. Tastants enter the body and are dissolved in saliva. Taste cells are located within taste buds, which are found on three of the four types of papillae in the mouth.

Regarding olfaction, there are many thousands of odorants, but humans detect only about 10,000. Like taste receptors, olfactory receptors are each responsive to only one odorant. Odorants dissolve in nasal mucosa, where they excite their corresponding olfactory sensory cells. When these cells detect an odorant, they send their signals to the main olfactory bulb and then to other locations in the brain, including the olfactory cortex.

## Review Questions

### Exercise:

**Problem:** Which of the following has the fewest taste receptors?

- a. fungiform papillae
- b. circumvallate papillae
- c. foliate papillae
- d. filiform papillae

---

### Solution:

D

### Exercise:

**Problem:**

How many different taste molecules do taste cells each detect?

- a. one
- b. five
- c. ten
- d. It depends on the spot on the tongue

---

**Solution:**

A

**Exercise:**

**Problem:** Salty foods activate the taste cells by \_\_\_\_\_.

- a. exciting the taste cell directly
- b. causing hydrogen ions to enter the cell
- c. causing sodium channels to close
- d. binding directly to the receptors

---

**Solution:**

A

**Exercise:****Problem:**

All sensory signals except \_\_\_\_\_ travel to the \_\_\_\_\_ in the brain before the cerebral cortex.

- a. vision; thalamus
- b. olfaction; thalamus
- c. vision; cranial nerves
- d. olfaction; cranial nerves

---

**Solution:**

B

**Free Response****Exercise:****Problem:**

From the perspective of the recipient of the signal, in what ways do pheromones differ from other odorants?

---

**Solution:**

Pheromones may not be consciously perceived, and pheromones can have direct physiological and behavioral effects on their recipients.

**Exercise:****Problem:**

What might be the effect on an animal of not being able to perceive taste?

---

**Solution:**

The animal might not be able to recognize the differences in food sources and thus might not be able to discriminate between spoiled food and safe food or between foods that contain necessary nutrients, such as proteins, and foods that do not.

**Glossary**

bipolar neuron

neuron with two processes from the cell body, typically in opposite directions

glomerulus

in the olfactory bulb, one of the two neural clusters that receives signals from one type of olfactory receptor

gustation

sense of taste

odorant

airborne molecule that stimulates an olfactory receptor

olfaction

sense of smell

olfactory bulb

neural structure in the vertebrate brain that receives signals from olfactory receptors

olfactory epithelium

specialized tissue in the nasal cavity where olfactory receptors are located

olfactory receptor

dendrite of a specialized neuron

papilla

one of the small bump-like projections from the tongue

pheromone

substance released by an animal that can affect the physiology or behavior of other animals

tastant

food molecule that stimulates gustatory receptors

taste bud

clusters of taste cells

umami

one of the five basic tastes, which is described as “savory” and which may be largely the taste of L-glutamate



## Hearing and Vestibular Sensation

By the end of this section, you will be able to:

- Describe the relationship of amplitude and frequency of a sound wave to attributes of sound
- Trace the path of sound through the auditory system to the site of transduction of sound
- Identify the structures of the vestibular system that respond to gravity

**Audition**, or hearing, is important to humans and to other animals for many different interactions. It enables an organism to detect and receive information about danger, such as an approaching predator, and to participate in communal exchanges like those concerning territories or mating. On the other hand, although it is physically linked to the auditory system, the vestibular system is not involved in hearing. Instead, an animal's vestibular system detects its own movement, both linear and angular acceleration and deceleration, and balance.

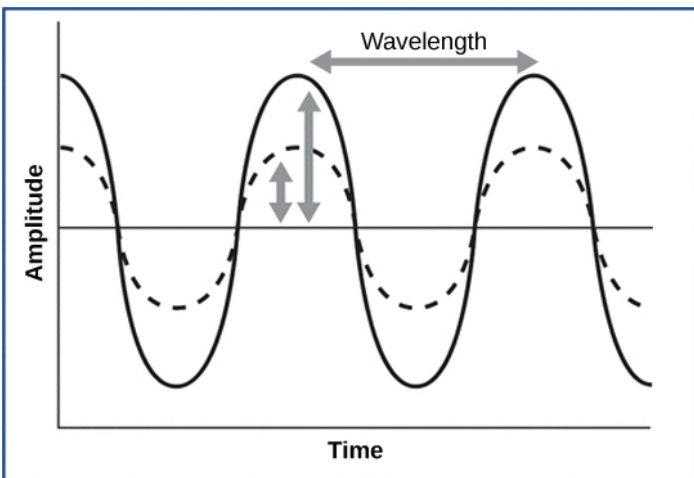
## Sound

Auditory stimuli are sound waves, which are mechanical, pressure waves that move through a medium, such as air or water. There are no sound waves in a vacuum since there are no air molecules to move in waves. The speed of sound waves differs, based on altitude, temperature, and medium, but at sea level and a temperature of 20° C (68° F), sound waves travel in the air at about 343 meters per second.

As is true for all waves, there are four main characteristics of a sound wave: frequency, wavelength, period, and amplitude. Frequency is the number of waves per unit of time, and in sound is heard as pitch. High-frequency ( $\geq 15,000\text{Hz}$ ) sounds are higher-pitched (short wavelength) than low-frequency (long wavelengths;  $\leq 100\text{Hz}$ ) sounds. Frequency is measured in cycles per second, and for sound, the most commonly used unit is hertz (Hz), or cycles per second. Most humans can perceive sounds with frequencies between 30 and 20,000 Hz. Women are typically better at hearing high frequencies, but everyone's ability to hear high frequencies decreases with age. Dogs detect up to about 40,000 Hz; cats, 60,000 Hz;

bats, 100,000 Hz; and dolphins 150,000 Hz, and American shad (*Alosa sapidissima*), a fish, can hear 180,000 Hz. Those frequencies above the human range are called **ultrasound**.

Amplitude, or the dimension of a wave from peak to trough, in sound is heard as volume and is illustrated in [\[link\]](#). The sound waves of louder sounds have greater amplitude than those of softer sounds. For sound, volume is measured in decibels (dB). The softest sound that a human can hear is the zero point. Humans speak normally at 60 decibels.

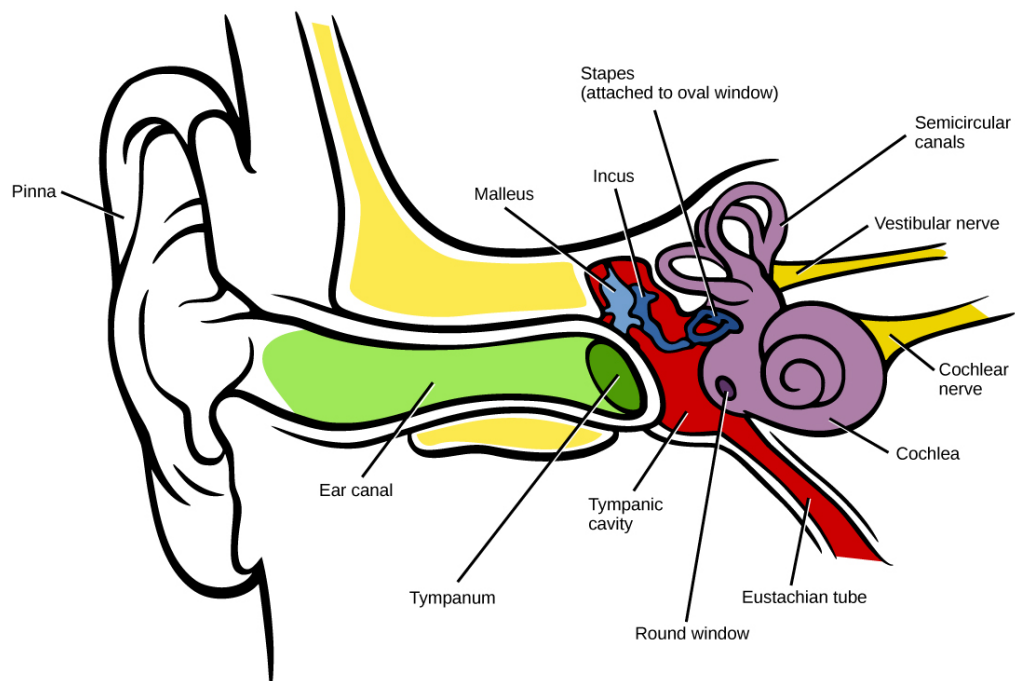


For sound waves, wavelength corresponds to pitch. Amplitude of the wave corresponds to volume. The sound wave shown with a dashed line is softer in volume than the sound wave shown with a solid line. (credit: NIH)

## Reception of Sound

In mammals, sound waves are collected by the external, cartilaginous part of the ear called the **pinna**, then travel through the auditory canal and cause

vibration of the thin diaphragm called the **tympanum** or ear drum, the innermost part of the **outer ear** (illustrated in [\[link\]](#)). Interior to the tympanum is the **middle ear**. The middle ear holds three small bones called the **ossicles**, which transfer energy from the moving tympanum to the inner ear. The three ossicles are the **malleus** (also known as the hammer), the **incus** (the anvil), and **stapes** (the stirrup). The aptly named stapes looks very much like a stirrup. The three ossicles are unique to mammals, and each plays a role in hearing. The malleus attaches at three points to the interior surface of the tympanic membrane. The incus attaches the malleus to the stapes. In humans, the stapes is not long enough to reach the tympanum. If we did not have the malleus and the incus, then the vibrations of the tympanum would never reach the inner ear. These bones also function to collect force and amplify sounds. The ear ossicles are homologous to bones in a fish mouth: the bones that support gills in fish are thought to be adapted for use in the vertebrate ear over evolutionary time. Many animals (frogs, reptiles, and birds, for example) use the stapes of the middle ear to transmit vibrations to the middle ear.



Sound travels through the outer ear to the middle ear, which is bounded on its exterior by the tympanic membrane. The

middle ear contains three bones called ossicles that transfer the sound wave to the oval window, the exterior boundary of the inner ear. The organ of Corti, which is the organ of sound transduction, lies inside the cochlea. (credit: modification of work by Lars Chittka, Axel Brockmann)

## Transduction of Sound

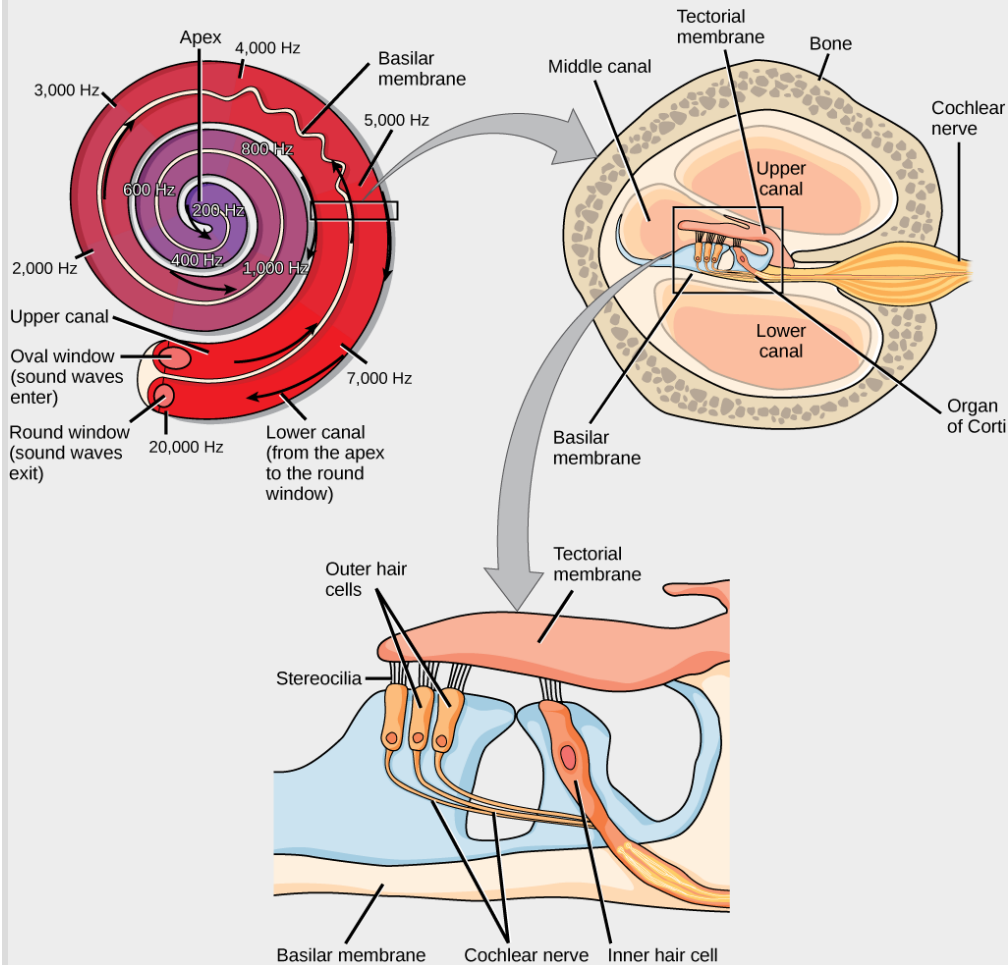
Vibrating objects, such as vocal cords, create sound waves or pressure waves in the air. When these pressure waves reach the ear, the ear transduces this mechanical stimulus (pressure wave) into a nerve impulse (electrical signal) that the brain perceives as sound. The pressure waves strike the tympanum, causing it to vibrate. The mechanical energy from the moving tympanum transmits the vibrations to the three bones of the middle ear. The stapes transmits the vibrations to a thin diaphragm called the **oval window**, which is the outermost structure of the **inner ear**. The structures of the inner ear are found in the **labyrinth**, a bony, hollow structure that is the most interior portion of the ear. Here, the energy from the sound wave is transferred from the stapes through the flexible oval window and to the fluid of the cochlea. The vibrations of the oval window create pressure waves in the fluid (perilymph) inside the cochlea. The **cochlea** is a whorled structure, like the shell of a snail, and it contains receptors for transduction of the mechanical wave into an electrical signal (as illustrated in [\[link\]](#)). Inside the cochlea, the **basilar membrane** is a mechanical analyzer that runs the length of the cochlea, curling toward the cochlea's center.

The mechanical properties of the basilar membrane change along its length, such that it is thicker, tauter, and narrower at the outside of the whorl (where the cochlea is largest), and thinner, floppier, and broader toward the apex, or center, of the whorl (where the cochlea is smallest). Different regions of the basilar membrane vibrate according to the frequency of the sound wave conducted through the fluid in the cochlea. For these reasons, the fluid-filled cochlea detects different wave frequencies (pitches) at different regions of the membrane. When the sound waves in the cochlear

fluid contact the basilar membrane, it flexes back and forth in a wave-like fashion. Above the basilar membrane is the **tectorial membrane**.

**Note:**

**Art Connection**



In the human ear, sound waves cause the stapes to press against the oval window. Vibrations travel up the fluid-filled interior of the cochlea. The basilar membrane that lines the cochlea gets continuously thinner toward the apex of the cochlea. Different thicknesses of membrane vibrate in response to different frequencies of sound. Sound waves then exit through the round window. In the cross section of the cochlea (top right figure), note that in

addition to the upper canal and lower canal, the cochlea also has a middle canal. The organ of Corti (bottom image) is the site of sound transduction. Movement of stereocilia on hair cells results in an action potential that travels along the auditory nerve.

Cochlear implants can restore hearing in people who have a nonfunctional cochlear. The implant consists of a microphone that picks up sound. A speech processor selects sounds in the range of human speech, and a transmitter converts these sounds to electrical impulses, which are then sent to the auditory nerve. Which of the following types of hearing loss would not be restored by a cochlear implant?

- a. Hearing loss resulting from absence or loss of hair cells in the organ of Corti.
- b. Hearing loss resulting from an abnormal auditory nerve.
- c. Hearing loss resulting from fracture of the cochlea.
- d. Hearing loss resulting from damage to bones of the middle ear.

The site of transduction is in the **organ of Corti** (spiral organ). It is composed of hair cells held in place above the basilar membrane like flowers projecting up from soil, with their exposed short, hair-like **stereocilia** contacting or embedded in the tectorial membrane above them. The inner hair cells are the primary auditory receptors and exist in a single row, numbering approximately 3,500. The stereocilia from inner hair cells extend into small dimples on the tectorial membrane's lower surface. The outer hair cells are arranged in three or four rows. They number approximately 12,000, and they function to fine tune incoming sound waves. The longer stereocilia that project from the outer hair cells actually attach to the tectorial membrane. All of the stereocilia are mechanoreceptors, and when bent by vibrations they respond by opening a gated ion channel (refer to [\[link\]](#)). As a result, the hair cell membrane is depolarized, and a signal is transmitted to the cochlear nerve. Intensity

(volume) of sound is determined by how many hair cells at a particular location are stimulated.

The hair cells are arranged on the basilar membrane in an orderly way. The basilar membrane vibrates in different regions, according to the frequency of the sound waves impinging on it. Likewise, the hair cells that lay above it are most sensitive to a specific frequency of sound waves. Hair cells can respond to a small range of similar frequencies, but they require stimulation of greater intensity to fire at frequencies outside of their optimal range. The difference in response frequency between adjacent inner hair cells is about 0.2 percent. Compare that to adjacent piano strings, which are about six percent different. Place theory, which is the model for how biologists think pitch detection works in the human ear, states that high frequency sounds selectively vibrate the basilar membrane of the inner ear near the entrance port (the oval window). Lower frequencies travel farther along the membrane before causing appreciable excitation of the membrane. The basic pitch-determining mechanism is based on the location along the membrane where the hair cells are stimulated. The place theory is the first step toward an understanding of pitch perception. Considering the extreme pitch sensitivity of the human ear, it is thought that there must be some auditory “sharpening” mechanism to enhance the pitch resolution.

When sound waves produce fluid waves inside the cochlea, the basilar membrane flexes, bending the stereocilia that attach to the tectorial membrane. Their bending results in action potentials in the hair cells, and auditory information travels along the neural endings of the bipolar neurons of the hair cells (collectively, the auditory nerve) to the brain. When the hairs bend, they release an excitatory neurotransmitter at a synapse with a sensory neuron, which then conducts action potentials to the central nervous system. The cochlear branch of the vestibulocochlear cranial nerve sends information on hearing. The auditory system is very refined, and there is some modulation or “sharpening” built in. The brain can send signals back to the cochlea, resulting in a change of length in the outer hair cells, sharpening or dampening the hair cells’ response to certain frequencies.

**Note:**

## Link to Learning



Watch an [animation](#) of sound entering the outer ear, moving through the ear structure, stimulating cochlear nerve impulses, and eventually sending signals to the temporal lobe.

## Higher Processing

The inner hair cells are most important for conveying auditory information to the brain. About 90 percent of the afferent neurons carry information from inner hair cells, with each hair cell synapsing with 10 or so neurons. Outer hair cells connect to only 10 percent of the afferent neurons, and each afferent neuron innervates many hair cells. The afferent, bipolar neurons that convey auditory information travel from the cochlea to the medulla, through the pons and midbrain in the brainstem, finally reaching the primary auditory cortex in the temporal lobe.

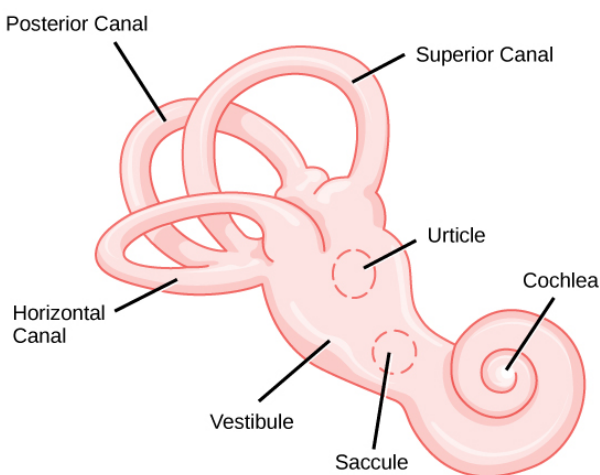
## Vestibular Information

The stimuli associated with the vestibular system are linear acceleration (gravity) and angular acceleration and deceleration. Gravity, acceleration, and deceleration are detected by evaluating the inertia on receptive cells in the vestibular system. Gravity is detected through head position. Angular acceleration and deceleration are expressed through turning or tilting of the head.

The vestibular system has some similarities with the auditory system. It utilizes hair cells just like the auditory system, but it excites them in different ways. There are five vestibular receptor organs in the inner ear: the



utricle, the saccule, and three semicircular canals. Together, they make up what's known as the vestibular labyrinth that is shown in [\[link\]](#). The utricle and saccule respond to acceleration in a straight line, such as gravity. The roughly 30,000 hair cells in the utricle and 16,000 hair cells in the saccule lie below a gelatinous layer, with their stereocilia projecting into the gelatin. Embedded in this gelatin are calcium carbonate crystals—like tiny rocks. When the head is tilted, the crystals continue to be pulled straight down by gravity, but the new angle of the head causes the gelatin to shift, thereby bending the stereocilia. The bending of the stereocilia stimulates the neurons, and they signal to the brain that the head is tilted, allowing the maintenance of balance. It is the vestibular branch of the vestibulocochlear cranial nerve that deals with balance.



The structure of the vestibular labyrinth is shown. (credit: modification of work by NIH)

The fluid-filled **semicircular canals** are tubular loops set at oblique angles. They are arranged in three spatial planes. The base of each canal has a swelling that contains a cluster of hair cells. The hairs project into a gelatinous cap called the cupula and monitor angular acceleration and deceleration from rotation. They would be stimulated by driving your car

around a corner, turning your head, or falling forward. One canal lies horizontally, while the other two lie at about 45 degree angles to the horizontal axis, as illustrated in [\[link\]](#). When the brain processes input from all three canals together, it can detect angular acceleration or deceleration in three dimensions. When the head turns, the fluid in the canals shifts, thereby bending stereocilia and sending signals to the brain. Upon cessation accelerating or decelerating—or just moving—the movement of the fluid within the canals slows or stops. For example, imagine holding a glass of water. When moving forward, water may splash backwards onto the hand, and when motion has stopped, water may splash forward onto the fingers. While in motion, the water settles in the glass and does not splash. Note that the canals are not sensitive to velocity itself, but to changes in velocity, so moving forward at 60mph with your eyes closed would not give the sensation of movement, but suddenly accelerating or braking would stimulate the receptors.

## Higher Processing

Hair cells from the utricle, saccule, and semicircular canals also communicate through bipolar neurons to the cochlear nucleus in the medulla. Cochlear neurons send descending projections to the spinal cord and ascending projections to the pons, thalamus, and cerebellum. Connections to the cerebellum are important for coordinated movements. There are also projections to the temporal cortex, which account for feelings of dizziness; projections to autonomic nervous system areas in the brainstem, which account for motion sickness; and projections to the primary somatosensory cortex, which monitors subjective measurements of the external world and self-movement. People with lesions in the vestibular area of the somatosensory cortex see vertical objects in the world as being tilted. Finally, the vestibular signals project to certain optic muscles to coordinate eye and head movements.

### **Note:**

Link to Learning



Click through this [interactive tutorial](#) to review the parts of the ear and how they function to process sound.

## Section Summary

Audition is important for territory defense, predation, predator defense, and communal exchanges. The vestibular system, which is not auditory, detects linear acceleration and angular acceleration and deceleration. Both the auditory system and vestibular system use hair cells as their receptors.

Auditory stimuli are sound waves. The sound wave energy reaches the outer ear (pinna, canal, tympanum), and vibrations of the tympanum send the energy to the middle ear. The middle ear bones shift and the stapes transfers mechanical energy to the oval window of the fluid-filled inner ear cochlea. Once in the cochlea, the energy causes the basilar membrane to flex, thereby bending the stereocilia on receptor hair cells. This activates the receptors, which send their auditory neural signals to the brain.

The vestibular system has five parts that work together to provide the sense of direction, thus helping to maintain balance. The utricle and saccule measure head orientation: their calcium carbonate crystals shift when the head is tilted, thereby activating hair cells. The semicircular canals work similarly, such that when the head is turned, the fluid in the canals bends stereocilia on hair cells. The vestibular hair cells also send signals to the thalamus and to somatosensory cortex, but also to the cerebellum, the structure above the brainstem that plays a large role in timing and coordination of movement.

## Art Connections

**Exercise:****Problem:**

[\[link\]](#) Cochlear implants can restore hearing in people who have a nonfunctional cochlear. The implant consists of a microphone that picks up sound. A speech processor selects sounds in the range of human speech, and a transmitter converts these sounds to electrical impulses, which are then sent to the auditory nerve. Which of the following types of hearing loss would not be restored by a cochlear implant?

- a. Hearing loss resulting from absence or loss of hair cells in the organ of Corti.
- b. Hearing loss resulting from an abnormal auditory nerve.
- c. Hearing loss resulting from fracture of the cochlea.
- d. Hearing loss resulting from damage to bones of the middle ear.

---

**Solution:**

[\[link\]](#) B

**Review Questions****Exercise:****Problem:**

In sound, pitch is measured in \_\_\_\_\_, and volume is measured in \_\_\_\_\_.

- a. nanometers (nm); decibels (dB)
- b. decibels (dB); nanometers (nm)
- c. decibels (dB); hertz (Hz)
- d. hertz (Hz); decibels (dB)

---

**Solution:**

D

**Exercise:**

**Problem:** Auditory hair cells are indirectly anchored to the \_\_\_\_\_.

- a. basilar membrane
- b. oval window
- c. tectorial membrane
- d. ossicles

---

**Solution:**

A

**Exercise:**

**Problem:**

Which of the following are found both in the auditory system and the vestibular system?

- a. basilar membrane
- b. hair cells
- c. semicircular canals
- d. ossicles

---

**Solution:**

B

**Free Response**

**Exercise:**

**Problem:**

How would a rise in altitude likely affect the speed of a sound transmitted through air? Why?

---

**Solution:**

The sound would slow down, because it is transmitted through the particles (gas) and there are fewer particles (lower density) at higher altitudes.

**Exercise:****Problem:**

How might being in a place with less gravity than Earth has (such as Earth's moon) affect vestibular sensation, and why?

---

**Solution:**

Because vestibular sensation relies on gravity's effects on tiny crystals in the inner ear, a situation of reduced gravity would likely impair vestibular sensation.

**Glossary**

audition

sense of hearing

basilar membrane

stiff structure in the cochlea that indirectly anchors auditory receptors

cochlea

whorled structure that contains receptors for transduction of the mechanical wave into an electrical signal

incus

(also, anvil) second of the three bones of the middle ear

inner ear

innermost part of the ear; consists of the cochlea and the vestibular system

labyrinth

bony, hollow structure that is the most internal part of the ear; contains the sites of transduction of auditory and vestibular information

malleus

(also, hammer) first of the three bones of the middle ear

middle ear

part of the hearing apparatus that functions to transfer energy from the tympanum to the oval window of the inner ear

organ of Corti

in the basilar membrane, the site of the transduction of sound, a mechanical wave, to a neural signal

ossicle

one of the three bones of the middle ear

outer ear

part of the ear that consists of the pinna, ear canal, and tympanum and which conducts sound waves into the middle ear

oval window

thin diaphragm between the middle and inner ears that receives sound waves from contact with the stapes bone of the middle ear

pinna

cartilaginous outer ear

semicircular canal

one of three half-circular, fluid-filled tubes in the vestibular labyrinth that monitors angular acceleration and deceleration

stapes

(also, stirrup) third of the three bones of the middle ear

stereocilia

in the auditory system, hair-like projections from hair cells that help detect sound waves

tectorial membrane

cochlear structure that lies above the hair cells and participates in the transduction of sound at the hair cells

tympanum

(also, tympanic membrane or ear drum) thin diaphragm between the outer and middle ears

ultrasound

sound frequencies above the human detectable ceiling of approximately 20,000 Hz



## Vision

By the end of this section, you will be able to:

- Explain how electromagnetic waves differs from sound waves
- Trace the path of light through the eye to the point of the optic nerve
- Explain tonic activity as it is manifested in photoreceptors in the retina

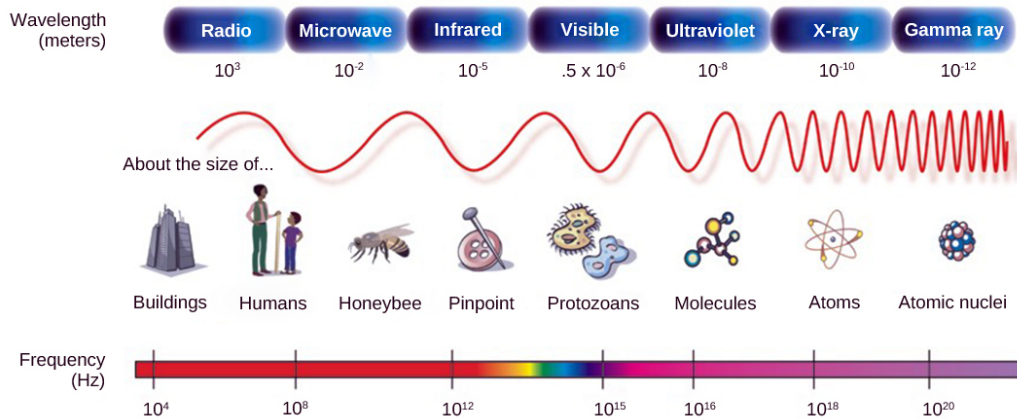
**Vision** is the ability to detect light patterns from the outside environment and interpret them into images. Animals are bombarded with sensory information, and the sheer volume of visual information can be problematic. Fortunately, the visual systems of species have evolved to attend to the most-important stimuli. The importance of vision to humans is further substantiated by the fact that about one-third of the human cerebral cortex is dedicated to analyzing and perceiving visual information.

## Light

As with auditory stimuli, light travels in waves. The compression waves that compose sound must travel in a medium—a gas, a liquid, or a solid. In contrast, light is composed of electromagnetic waves and needs no medium; light can travel in a vacuum ([\[link\]](#)). The behavior of light can be discussed in terms of the behavior of waves and also in terms of the behavior of the fundamental unit of light—a packet of electromagnetic radiation called a photon. A glance at the electromagnetic spectrum shows that visible light for humans is just a small slice of the entire spectrum, which includes radiation that we cannot see as light because it is below the frequency of visible red light and above the frequency of visible violet light.

Certain variables are important when discussing perception of light. Wavelength (which varies inversely with frequency) manifests itself as hue. Light at the red end of the visible spectrum has longer wavelengths (and is lower frequency), while light at the violet end has shorter wavelengths (and is higher frequency). The wavelength of light is expressed in nanometers (nm); one nanometer is one billionth of a meter. Humans perceive light that ranges between approximately 380 nm and 740 nm. Some other animals, though, can detect wavelengths outside of the human range. For example,

bees see near-ultraviolet light in order to locate nectar guides on flowers, and some non-avian reptiles sense infrared light (heat that prey gives off).



In the electromagnetic spectrum, visible light lies between 380 nm and 740 nm. (credit: modification of work by NASA)

Wave amplitude is perceived as luminous intensity, or brightness. The standard unit of intensity of light is the **candela**, which is approximately the luminous intensity of a one common candle.

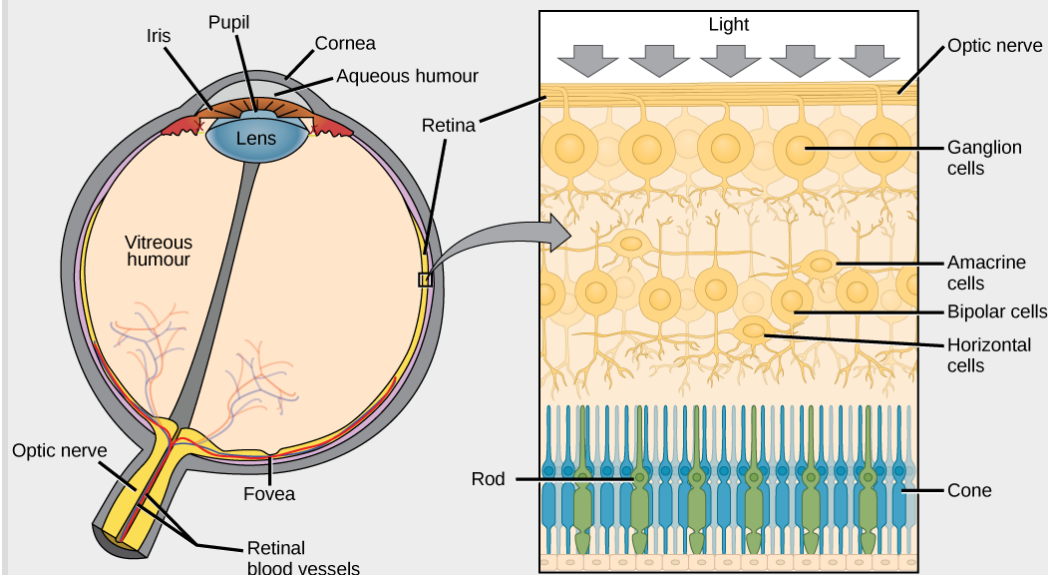
Light waves travel 299,792 km per second in a vacuum, (and somewhat slower in various media such as air and water), and those waves arrive at the eye as long (red), medium (green), and short (blue) waves. What is termed “white light” is light that is perceived as white by the human eye. This effect is produced by light that stimulates equally the color receptors in the human eye. The apparent color of an object is the color (or colors) that the object reflects. Thus a red object reflects the red wavelengths in mixed (white) light and absorbs all other wavelengths of light.

## Anatomy of the Eye

The photoreceptive cells of the eye, where transduction of light to nervous impulses occurs, are located in the **retina** (shown in [\[link\]](#)) on the inner surface of the back of the eye. But light does not impinge on the retina unaltered. It passes through other layers that process it so that it can be interpreted by the retina ([\[link\]](#)**b**). The **cornea**, the front transparent layer of the eye, and the crystalline **lens**, a transparent convex structure behind the cornea, both refract (bend) light to focus the image on the retina. The **iris**, which is conspicuous as the colored part of the eye, is a circular muscular ring lying between the lens and cornea that regulates the amount of light entering the eye. In conditions of high ambient light, the iris contracts, reducing the size of the pupil at its center. In conditions of low light, the iris relaxes and the pupil enlarges.

**Note:**

**Art Connection**



(a) The human eye is shown in cross section. (b) A blowup shows the layers of the retina.

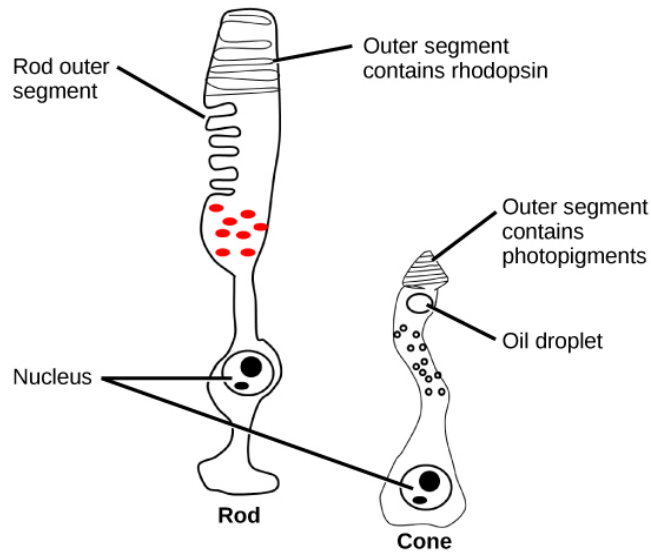
Which of the following statements about the human eye is false?

- a. Rods detect color, while cones detect only shades of gray.

- b. When light enters the retina, it passes the ganglion cells and bipolar cells before reaching photoreceptors at the rear of the eye.
- c. The iris adjusts the amount of light coming into the eye.
- d. The cornea is a protective layer on the front of the eye.

The main function of the lens is to focus light on the retina and fovea centralis. The lens is dynamic, focusing and re-focusing light as the eye rests on near and far objects in the visual field. The lens is operated by muscles that stretch it flat or allow it to thicken, changing the focal length of light coming through it to focus it sharply on the retina. With age comes the loss of the flexibility of the lens, and a form of farsightedness called **presbyopia** results. Presbyopia occurs because the image focuses behind the retina. Presbyopia is a deficit similar to a different type of farsightedness called **hyperopia** caused by an eyeball that is too short. For both defects, images in the distance are clear but images nearby are blurry. **Myopia** (nearsightedness) occurs when an eyeball is elongated and the image focus falls in front of the retina. In this case, images in the distance are blurry but images nearby are clear.

There are two types of photoreceptors in the retina: **rods** and **cones**, named for their general appearance as illustrated in [\[link\]](#). Rods are strongly photosensitive and are located in the outer edges of the retina. They detect dim light and are used primarily for peripheral and nighttime vision. Cones are weakly photosensitive and are located near the center of the retina. They respond to bright light, and their primary role is in daytime, color vision.



Rods and cones are photoreceptors in the retina. Rods respond in low light and can detect only shades of gray. Cones respond in intense light and are responsible for color vision. (credit: modification of work by Piotr Sliwa)

The **fovea** is the region in the center back of the eye that is responsible for acute vision. The fovea has a high density of cones. When you bring your gaze to an object to examine it intently in bright light, the eyes orient so that the object's image falls on the fovea. However, when looking at a star in the night sky or other object in dim light, the object can be better viewed by the peripheral vision because it is the rods at the edges of the retina, rather than the cones at the center, that operate better in low light. In humans, cones far outnumber rods in the fovea.

**Note:**

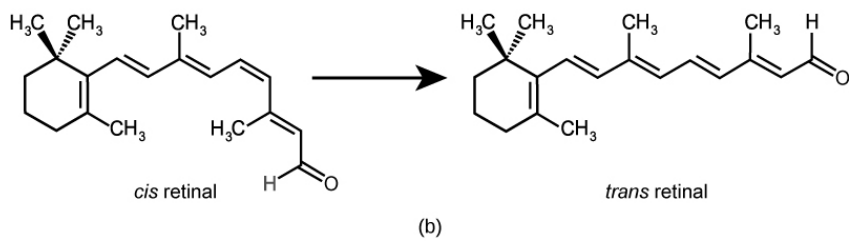
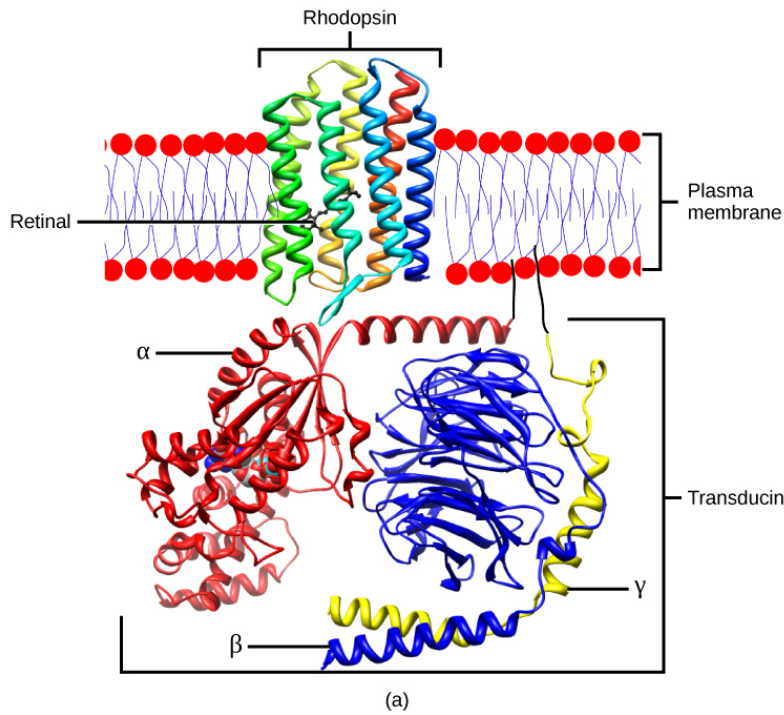
[Link to Learning](#)



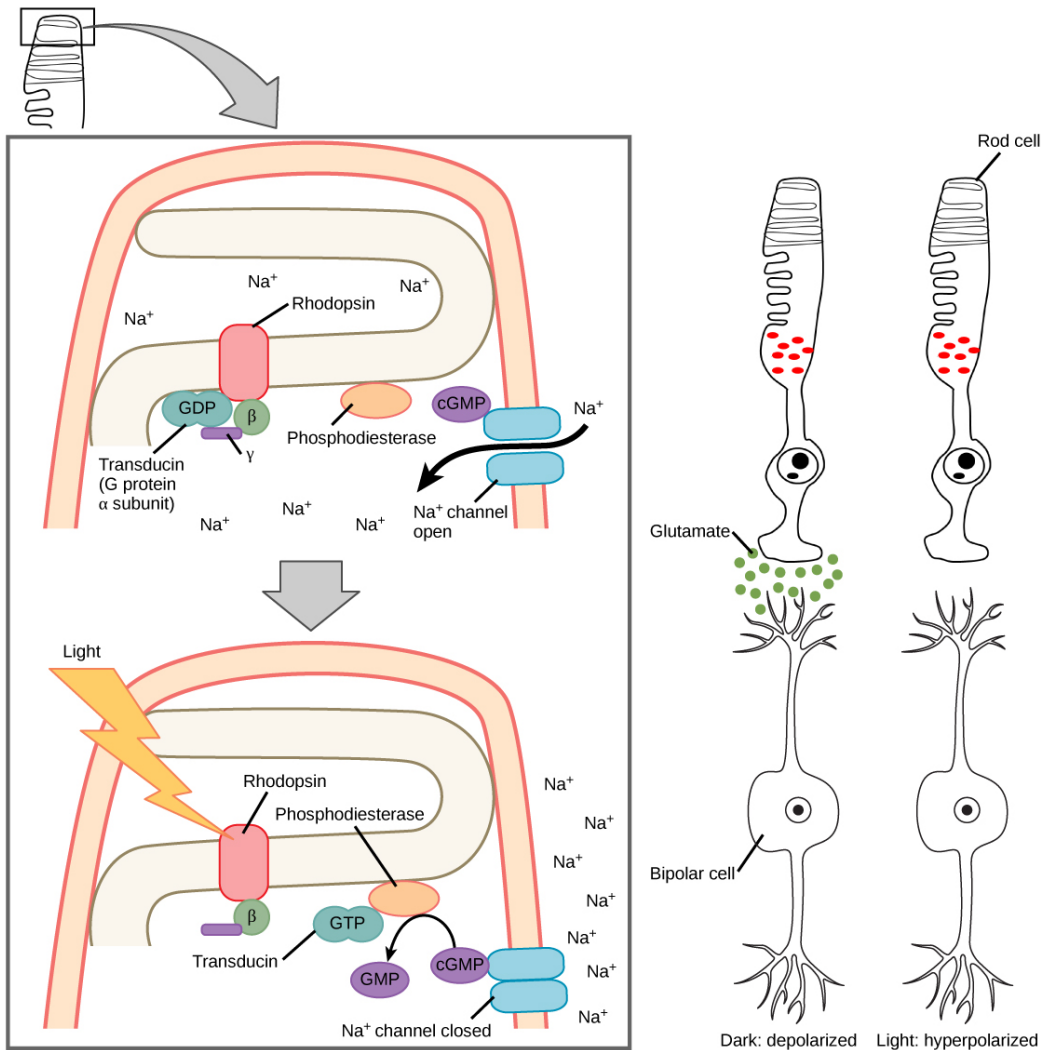
Review the [anatomical structure](#) of the eye, clicking on each part to practice identification.

## Transduction of Light

The rods and cones are the site of transduction of light to a neural signal. Both rods and cones contain photopigments. In vertebrates, the main photopigment, **rhodopsin**, has two main parts ([link](#)): an opsin, which is a membrane protein (in the form of a cluster of  $\alpha$ -helices that span the membrane), and retinal—a molecule that absorbs light. When light hits a photoreceptor, it causes a shape change in the retinal, altering its structure from a bent (*cis*) form of the molecule to its linear (*trans*) isomer. This isomerization of retinal activates the rhodopsin, starting a cascade of events that ends with the closing of  $\text{Na}^+$  channels in the membrane of the photoreceptor. Thus, unlike most other sensory neurons (which become depolarized by exposure to a stimulus) visual receptors become hyperpolarized and thus driven away from threshold ([link](#)).



(a) Rhodopsin, the photoreceptor in vertebrates, has two parts: the trans-membrane protein opsin, and retinal. When light strikes retinal, it changes shape from (b) a *cis* to a *trans* form. The signal is passed to a G-protein called transducin, triggering a series of downstream events.



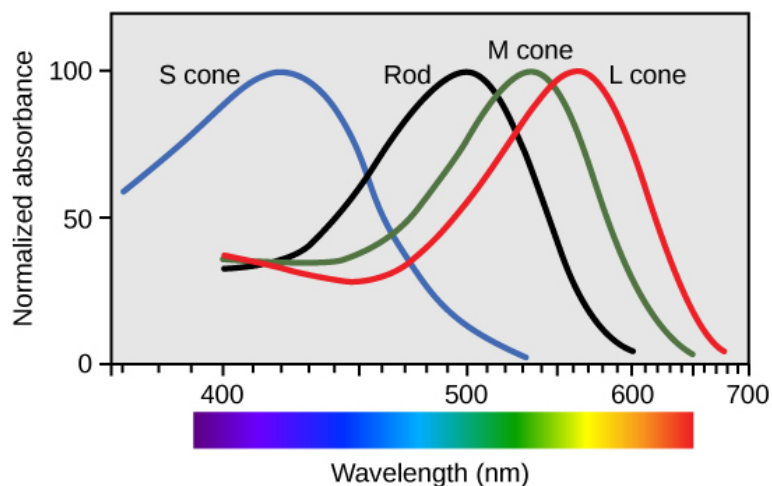
When light strikes rhodopsin, the G-protein transducin is activated, which in turn activates phosphodiesterase. Phosphodiesterase converts cGMP to GMP, thereby closing sodium channels. As a result, the membrane becomes hyperpolarized. The hyperpolarized membrane does not release glutamate to the bipolar cell.

## Trichromatic Coding



There are three types of cones (with different photopsins), and they differ in the wavelength to which they are most responsive, as shown in [\[link\]](#). Some cones are maximally responsive to short light waves of 420 nm, so they are called S cones (“S” for “short”); others respond maximally to waves of 530 nm (M cones, for “medium”); a third group responds maximally to light of longer wavelengths, at 560 nm (L, or “long” cones). With only one type of cone, color vision would not be possible, and a two-cone (dichromatic) system has limitations. Primates use a three-cone (trichromatic) system, resulting in full color vision.

The color we perceive is a result of the ratio of activity of our three types of cones. The colors of the visual spectrum, running from long-wavelength light to short, are red (700 nm), orange (600 nm), yellow (565 nm), green (497 nm), blue (470 nm), indigo (450 nm), and violet (425 nm). Humans have very sensitive perception of color and can distinguish about 500 levels of brightness, 200 different hues, and 20 steps of saturation, or about 2 million distinct colors.



Human rod cells and the different types of cone cells each have an optimal wavelength. However, there is considerable overlap in the wavelengths of light detected.

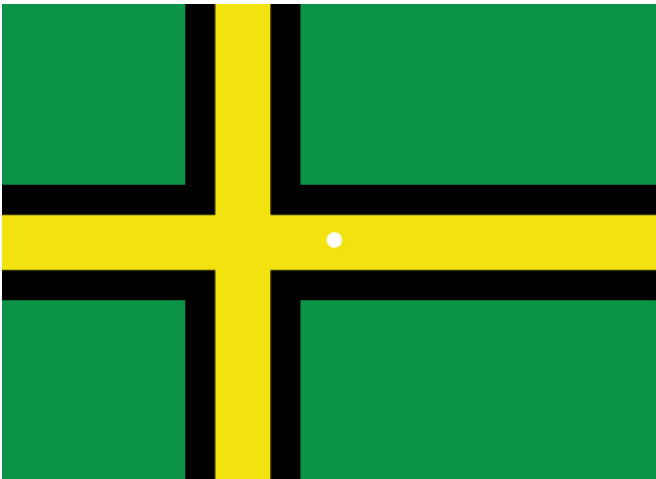
## Retinal Processing

Visual signals leave the cones and rods, travel to the bipolar cells, and then to ganglion cells. A large degree of processing of visual information occurs in the retina itself, before visual information is sent to the brain.

Photoreceptors in the retina continuously undergo **tonic activity**. That is, they are always slightly active even when not stimulated by light. In neurons that exhibit tonic activity, the absence of stimuli maintains a firing rate at a baseline; while some stimuli increase firing rate from the baseline, and other stimuli decrease firing rate. In the absence of light, the bipolar neurons that connect rods and cones to ganglion cells are continuously and actively inhibited by the rods and cones. Exposure of the retina to light hyperpolarizes the rods and cones and removes their inhibition of bipolar cells. The now active bipolar cells in turn stimulate the ganglion cells, which send action potentials along their axons (which leave the eye as the optic nerve). Thus, the visual system relies on change in retinal activity, rather than the absence or presence of activity, to encode visual signals for the brain. Sometimes horizontal cells carry signals from one rod or cone to other photoreceptors and to several bipolar cells. When a rod or cone stimulates a horizontal cell, the horizontal cell inhibits more distant photoreceptors and bipolar cells, creating lateral inhibition. This inhibition sharpens edges and enhances contrast in the images by making regions receiving light appear lighter and dark surroundings appear darker. Amacrine cells can distribute information from one bipolar cell to many ganglion cells.

You can demonstrate this using an easy demonstration to “trick” your retina and brain about the colors you are observing in your visual field. Look fixedly at [\[link\]](#) for about 45 seconds. Then quickly shift your gaze to a sheet of blank white paper or a white wall. You should see an afterimage of the Norwegian flag in its correct colors. At this point, close your eyes for a moment, then reopen them, looking again at the white paper or wall; the afterimage of the flag should continue to appear as red, white, and blue.

What causes this? According to an explanation called opponent process theory, as you gazed fixedly at the green, black, and yellow flag, your retinal ganglion cells that respond positively to green, black, and yellow increased their firing dramatically. When you shifted your gaze to the neutral white ground, these ganglion cells abruptly decreased their activity and the brain interpreted this abrupt downshift as if the ganglion cells were responding now to their “opponent” colors: red, white, and blue, respectively, in the visual field. Once the ganglion cells return to their baseline activity state, the false perception of color will disappear.



View this flag to understand how retinal processing works. Stare at the center of the flag (indicated by the white dot) for 45 seconds, and then quickly look at a white background, noticing how colors appear.

## Higher Processing

The myelinated axons of ganglion cells make up the optic nerves. Within the nerves, different axons carry different qualities of the visual signal.

Some axons constitute the magnocellular (big cell) pathway, which carries information about form, movement, depth, and differences in brightness. Other axons constitute the parvocellular (small cell) pathway, which carries information on color and fine detail. Some visual information projects directly back into the brain, while other information crosses to the opposite side of the brain. This crossing of optical pathways produces the distinctive optic chiasma (Greek, for “crossing”) found at the base of the brain and allows us to coordinate information from both eyes.

Once in the brain, visual information is processed in several places, and its routes reflect the complexity and importance of visual information to humans and other animals. One route takes the signals to the thalamus, which serves as the routing station for all incoming sensory impulses except olfaction. In the thalamus, the magnocellular and parvocellular distinctions remain intact, and there are different layers of the thalamus dedicated to each. When visual signals leave the thalamus, they travel to the primary visual cortex at the rear of the brain. From the visual cortex, the visual signals travel in two directions. One stream that projects to the parietal lobe, in the side of the brain, carries magnocellular (“where”) information. A second stream projects to the temporal lobe and carries both magnocellular (“where”) and parvocellular (“what”) information.

Another important visual route is a pathway from the retina to the **superior colliculus** in the midbrain, where eye movements are coordinated and integrated with auditory information. Finally, there is the pathway from the retina to the **suprachiasmatic nucleus** (SCN) of the hypothalamus. The SCN is a cluster of cells that is considered to be the body’s internal clock, which controls our **circadian** (day-long) cycle. The SCN sends information to the pineal gland, which is important in sleep/wake patterns and annual cycles.

**Note:**

[Link to Learning](#)



View this [interactive presentation](#) to review what you have learned about how vision functions.

## Section Summary

Vision is the only photo responsive sense. Visible light travels in waves and is a very small slice of the electromagnetic radiation spectrum. Light waves differ based on their frequency (wavelength = hue) and amplitude (intensity = brightness).

In the vertebrate retina, there are two types of light receptors (photoreceptors): cones and rods. Cones, which are the source of color vision, exist in three forms—L, M, and S—and they are differentially sensitive to different wavelengths. Cones are located in the retina, along with the dim-light, achromatic receptors (rods). Cones are found in the fovea, the central region of the retina, whereas rods are found in the peripheral regions of the retina.

Visual signals travel from the eye over the axons of retinal ganglion cells, which make up the optic nerves. Ganglion cells come in several versions. Some ganglion cell axons carry information on form, movement, depth, and brightness, while other axons carry information on color and fine detail. Visual information is sent to the superior colliculi in the midbrain, where coordination of eye movements and integration of auditory information takes place. Visual information is also sent to the suprachiasmatic nucleus (SCN) of the hypothalamus, which plays a role in the circadian cycle.

## Art Connections

**Exercise:****Problem:**

[\[link\]](#) Which of the following statements about the human eye is false?

- a. Rods detect color, while cones detect only shades of gray.
  - b. When light enters the retina, it passes the ganglion cells and bipolar cells before reaching photoreceptors at the rear of the eye.
  - c. The iris adjusts the amount of light coming into the eye.
  - d. The cornea is a protective layer on the front of the eye.
- 

**Solution:**

[\[link\]](#) A

**Review Questions****Exercise:**

**Problem:** Why do people over 55 often need reading glasses?

- a. Their cornea no longer focuses correctly.
  - b. Their lens no longer focuses correctly.
  - c. Their eyeball has elongated with age, causing images to focus in front of their retina.
  - d. Their retina has thinned with age, making vision more difficult.
- 

**Solution:**

B

**Exercise:**

**Problem:**

Why is it easier to see images at night using peripheral, rather than the central, vision?

- a. Cones are denser in the periphery of the retina.
- b. Bipolar cells are denser in the periphery of the retina.
- c. Rods are denser in the periphery of the retina.
- d. The optic nerve exits at the periphery of the retina.

---

**Solution:**

C

**Exercise:****Problem:**

A person catching a ball must coordinate her head and eyes. What part of the brain is helping to do this?

- a. hypothalamus
- b. pineal gland
- c. thalamus
- d. superior colliculus

---

**Solution:**

D

**Free Response****Exercise:**

**Problem:**

How could the pineal gland, the brain structure that plays a role in annual cycles, use visual information from the suprachiasmatic nucleus of the hypothalamus?

---

**Solution:**

The pineal gland could use length-of-day information to determine the time of year, for example. Day length is shorter in the winter than it is in the summer. For many animals and plants, photoperiod cues them to reproduce at a certain time of year.

**Exercise:****Problem:**

How is the relationship between photoreceptors and bipolar cells different from other sensory receptors and adjacent cells?

---

**Solution:**

The photoreceptors tonically inhibit the bipolar cells, and stimulation of the receptors turns this inhibition off, activating the bipolar cells.

**Glossary**

candela

(cd) unit of measurement of luminous intensity (brightness)

circadian

describes a time cycle about one day in length

cone

weakly photosensitive, chromatic, cone-shaped neuron in the fovea of the retina that detects bright light and is used in daytime color vision

cornea



transparent layer over the front of the eye that helps focus light waves

fovea

region in the center of the retina with a high density of photoreceptors and which is responsible for acute vision

hyperopia

(also, farsightedness) visual defect in which the image focus falls behind the retina, thereby making images in the distance clear, but close-up images blurry

iris

pigmented, circular muscle at the front of the eye that regulates the amount of light entering the eye

lens

transparent, convex structure behind the cornea that helps focus light waves on the retina

myopia

(also, nearsightedness) visual defect in which the image focus falls in front of the retina, thereby making images in the distance blurry, but close-up images clear

presbyopia

visual defect in which the image focus falls behind the retina, thereby making images in the distance clear, but close-up images blurry; caused by age-based changes in the lens

pupil

small opening through which light enters

retina

layer of photoreceptive and supporting cells on the inner surface of the back of the eye

rhodopsin

main photopigment in vertebrates

rod

strongly photosensitive, achromatic, cylindrical neuron in the outer edges of the retina that detects dim light and is used in peripheral and nighttime vision

superior colliculus

paired structure in the top of the midbrain, which manages eye movements and auditory integration

suprachiasmatic nucleus

cluster of cells in the hypothalamus that plays a role in the circadian cycle

tonic activity

in a neuron, slight continuous activity while at rest

vision

sense of sight

## Introduction

class="introduction"

The process of amphibian metamorphosis, as seen in the tadpole-to-frog stages shown here, is driven by hormones.

(credit "tadpole": modification of work by Brian Gratwicke)



An animal's endocrine system controls body processes through the production, secretion, and regulation of hormones, which serve as chemical "messengers" functioning in cellular and organ activity and, ultimately, maintaining the body's homeostasis. The endocrine system plays a role in growth, metabolism, and sexual development. In humans, common endocrine system diseases include thyroid disease and diabetes mellitus. In organisms that undergo metamorphosis, the process is controlled by the endocrine system. The transformation from tadpole to frog, for example, is complex and nuanced to adapt to specific environments and ecological circumstances.

## Endocrine Glands

By the end of this section, you will be able to:

- Describe the role of different glands in the endocrine system
- Explain how the different glands work together to maintain homeostasis

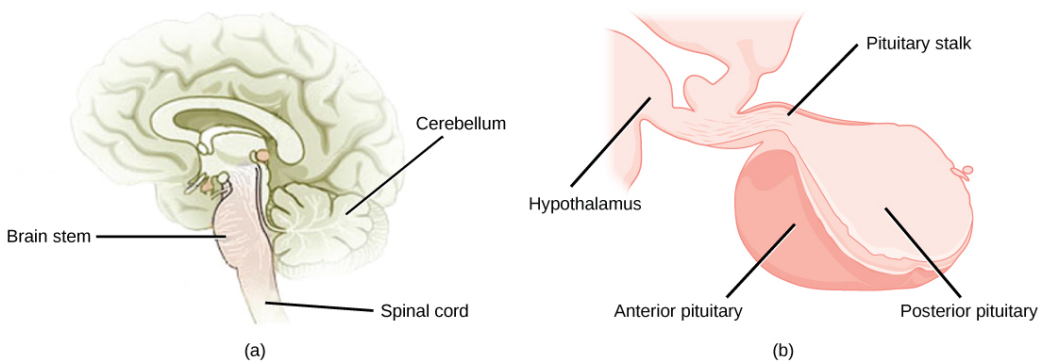
Both the endocrine and nervous systems use chemical signals to communicate and regulate the body's physiology. The endocrine system releases hormones that act on target cells to regulate development, growth, energy metabolism, reproduction, and many behaviors. The nervous system releases neurotransmitters or neurohormones that regulate neurons, muscle cells, and endocrine cells. Because the neurons can regulate the release of hormones, the nervous and endocrine systems work in a coordinated manner to regulate the body's physiology.

## Hypothalamic-Pituitary Axis

The **hypothalamus** in vertebrates integrates the endocrine and nervous systems. The hypothalamus is an endocrine organ located in the diencephalon of the brain. It receives input from the body and other brain areas and initiates endocrine responses to environmental changes. The hypothalamus acts as an endocrine organ, synthesizing hormones and transporting them along axons to the posterior pituitary gland. It synthesizes and secretes regulatory hormones that control the endocrine cells in the anterior pituitary gland. The hypothalamus contains autonomic centers that control endocrine cells in the adrenal medulla via neuronal control.

The **pituitary gland**, sometimes called the hypophysis or “master gland” is located at the base of the brain in the sella turcica, a groove of the sphenoid bone of the skull, illustrated in [\[link\]](#). It is attached to the hypothalamus via a stalk called the **pituitary stalk** (or infundibulum). The anterior portion of the pituitary gland is regulated by releasing or release-inhibiting hormones produced by the hypothalamus, and the posterior pituitary receives signals via neurosecretory cells to release hormones produced by the hypothalamus. The pituitary has two distinct regions—the anterior pituitary and the posterior pituitary—which between them secrete nine different

peptide or protein hormones. The posterior lobe of the pituitary gland contains axons of the hypothalamic neurons.



The pituitary gland is located at (a) the base of the brain and (b) connected to the hypothalamus by the pituitary stalk. (credit a: modification of work by NCI; credit b: modification of work by Gray's Anatomy)

## Anterior Pituitary

The **anterior pituitary** gland, or adenohypophysis, is surrounded by a capillary network that extends from the hypothalamus, down along the infundibulum, and to the anterior pituitary. This capillary network is a part of the **hypophyseal portal system** that carries substances from the hypothalamus to the anterior pituitary and hormones from the anterior pituitary into the circulatory system. A portal system carries blood from one capillary network to another; therefore, the hypophyseal portal system allows hormones produced by the hypothalamus to be carried directly to the anterior pituitary without first entering the circulatory system.

The anterior pituitary produces seven hormones: growth hormone (GH), prolactin (PRL), thyroid-stimulating hormone (TSH), melanin-stimulating hormone (MSH), adrenocorticotrophic hormone (ACTH), follicle-

stimulating hormone (FSH), and luteinizing hormone (LH). Anterior pituitary hormones are sometimes referred to as tropic hormones, because they control the functioning of other organs. While these hormones are produced by the anterior pituitary, their production is controlled by regulatory hormones produced by the hypothalamus. These regulatory hormones can be releasing hormones or inhibiting hormones, causing more or less of the anterior pituitary hormones to be secreted. These travel from the hypothalamus through the hypophyseal portal system to the anterior pituitary where they exert their effect. Negative feedback then regulates how much of these regulatory hormones are released and how much anterior pituitary hormone is secreted.

## Posterior Pituitary

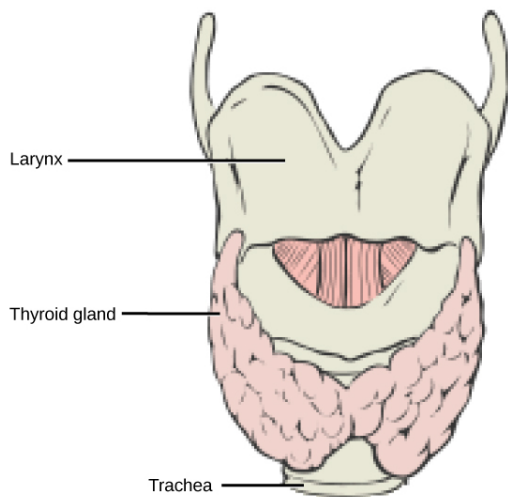
The **posterior pituitary** is significantly different in structure from the anterior pituitary. It is a part of the brain, extending down from the hypothalamus, and contains mostly nerve fibers and neuroglial cells, which support axons that extend from the hypothalamus to the posterior pituitary. The posterior pituitary and the infundibulum together are referred to as the neurohypophysis.

The hormones antidiuretic hormone (ADH), also known as vasopressin, and oxytocin are produced by neurons in the hypothalamus and transported within these axons along the infundibulum to the posterior pituitary. They are released into the circulatory system via neural signaling from the hypothalamus. These hormones are considered to be posterior pituitary hormones, even though they are produced by the hypothalamus, because that is where they are released into the circulatory system. The posterior pituitary itself does not produce hormones, but instead stores hormones produced by the hypothalamus and releases them into the blood stream.

## Thyroid Gland

The **thyroid gland** is located in the neck, just below the larynx and in front of the trachea, as shown in [\[link\]](#). It is a butterfly-shaped gland with two

lobes that are connected by the **isthmus**. It has a dark red color due to its extensive vascular system. When the thyroid swells due to dysfunction, it can be felt under the skin of the neck.



This illustration shows the location of the thyroid gland.

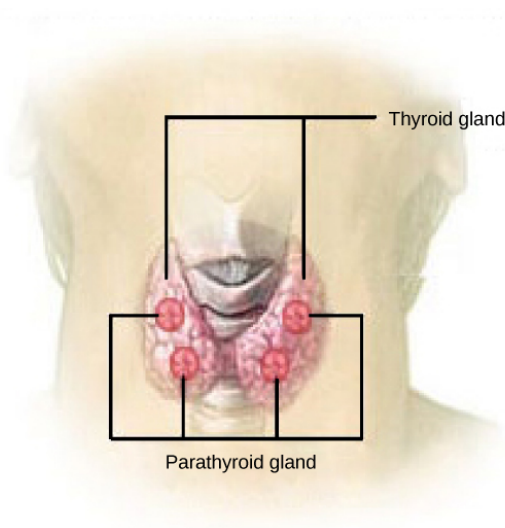
The thyroid gland is made up of many spherical thyroid follicles, which are lined with a simple cuboidal epithelium. These follicles contain a viscous fluid, called **colloid**, which stores the glycoprotein thyroglobulin, the precursor to the thyroid hormones. The follicles produce hormones that can be stored in the colloid or released into the surrounding capillary network for transport to the rest of the body via the circulatory system.

Thyroid follicle cells synthesize the hormone thyroxine, which is also known as  $T_4$  because it contains four atoms of iodine, and triiodothyronine, also known as  $T_3$  because it contains three atoms of iodine. Follicle cells are stimulated to release stored  $T_3$  and  $T_4$  by thyroid stimulating hormone (TSH), which is produced by the anterior pituitary. These thyroid hormones increase the rates of mitochondrial ATP production.

A third hormone, calcitonin, is produced by **parafollicular cells** of the thyroid either releasing hormones or inhibiting hormones. Calcitonin release is not controlled by TSH, but instead is released when calcium ion concentrations in the blood rise. Calcitonin functions to help regulate calcium concentrations in body fluids. It acts in the bones to inhibit osteoclast activity and in the kidneys to stimulate excretion of calcium. The combination of these two events lowers body fluid levels of calcium.

## Parathyroid Glands

Most people have four **parathyroid glands**; however, the number can vary from two to six. These glands are located on the posterior surface of the thyroid gland, as shown in [\[link\]](#). Normally, there is a superior gland and an inferior gland associated with each of the thyroid's two lobes. Each parathyroid gland is covered by connective tissue and contains many secretory cells that are associated with a capillary network.



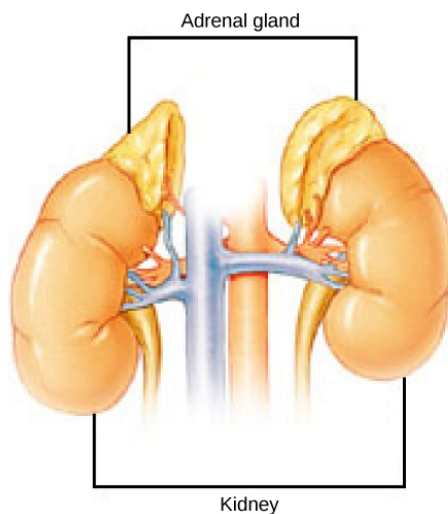
The parathyroid glands are located on the posterior of the thyroid gland. (credit: modification of work by NCI)



The parathyroid glands produce parathyroid hormone (PTH). PTH increases blood calcium concentrations when calcium ion levels fall below normal. PTH (1) enhances reabsorption of  $\text{Ca}^{2+}$  by the kidneys, (2) stimulates osteoclast activity and inhibits osteoblast activity, and (3) it stimulates synthesis and secretion of calcitriol by the kidneys, which enhances  $\text{Ca}^{2+}$  absorption by the digestive system. PTH is produced by chief cells of the parathyroid. PTH and calcitonin work in opposition to one another to maintain homeostatic  $\text{Ca}^{2+}$  levels in body fluids. Another type of cells, oxyphil cells, exist in the parathyroid but their function is not known. These hormones encourage bone growth, muscle mass, and blood cell formation in children and women.

## Adrenal Glands

The **adrenal glands** are associated with the kidneys; one gland is located on top of each kidney as illustrated in [\[link\]](#). The adrenal glands consist of an outer adrenal cortex and an inner adrenal medulla. These regions secrete different hormones.



The location of the adrenal glands on top of the kidneys

is shown. (credit:  
modification of work by  
NCI)

## **Adrenal Cortex**

The **adrenal cortex** is made up of layers of epithelial cells and associated capillary networks. These layers form three distinct regions: an outer zona glomerulosa that produces mineralocorticoids, a middle zona fasciculata that produces glucocorticoids, and an inner zona reticularis that produces androgens.

The main mineralocorticoid is aldosterone, which regulates the concentration of  $\text{Na}^+$  ions in urine, sweat, pancreas, and saliva. Aldosterone release from the adrenal cortex is stimulated by a decrease in blood concentrations of sodium ions, blood volume, or blood pressure, or by an increase in blood potassium levels.

The three main glucocorticoids are cortisol, corticosterone, and cortisone. The glucocorticoids stimulate the synthesis of glucose and gluconeogenesis (converting a non-carbohydrate to glucose) by liver cells and they promote the release of fatty acids from adipose tissue. These hormones increase blood glucose levels to maintain levels within a normal range between meals. These hormones are secreted in response to ACTH and levels are regulated by negative feedback.

Androgens are sex hormones that promote masculinity. They are produced in small amounts by the adrenal cortex in both males and females. They do not affect sexual characteristics and may supplement sex hormones released from the gonads.

## **Adrenal Medulla**

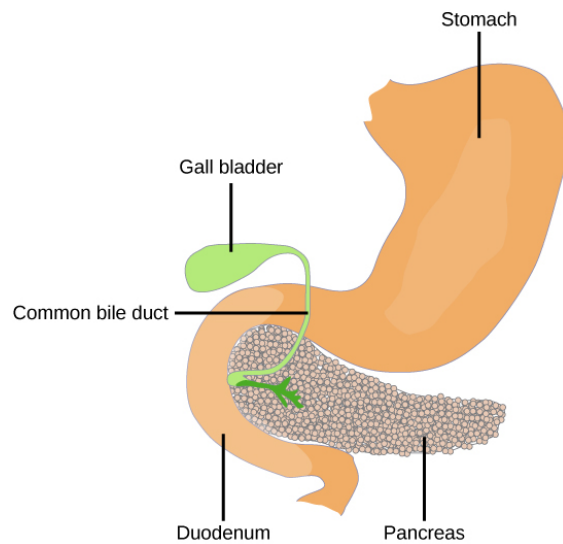
The **adrenal medulla** contains large, irregularly shaped cells that are closely associated with blood vessels. These cells are innervated by preganglionic autonomic nerve fibers from the central nervous system.

The adrenal medulla contains two types of secretory cells: one that produces epinephrine (adrenaline) and another that produces norepinephrine (noradrenaline). Epinephrine is the primary adrenal medulla hormone accounting for 75 to 80 percent of its secretions. Epinephrine and norepinephrine increase heart rate, breathing rate, cardiac muscle contractions, blood pressure, and blood glucose levels. They also accelerate the breakdown of glucose in skeletal muscles and stored fats in adipose tissue.

The release of epinephrine and norepinephrine is stimulated by neural impulses from the sympathetic nervous system. Secretion of these hormones is stimulated by acetylcholine release from preganglionic sympathetic fibers innervating the adrenal medulla. These neural impulses originate from the hypothalamus in response to stress to prepare the body for the fight-or-flight response.

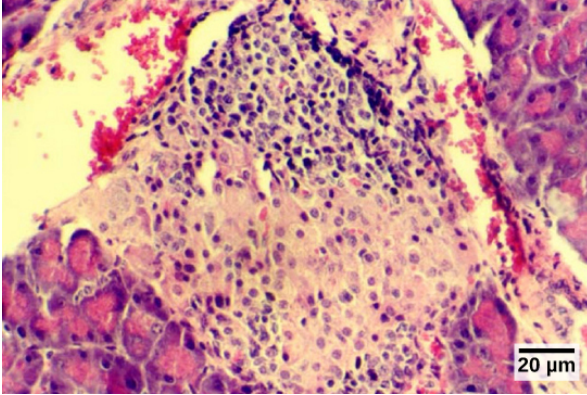
## **Pancreas**

The **pancreas**, illustrated in [\[link\]](#), is an elongated organ that is located between the stomach and the proximal portion of the small intestine. It contains both exocrine cells that excrete digestive enzymes and endocrine cells that release hormones. It is sometimes referred to as a heterocrine gland because it has both endocrine and exocrine functions.



The pancreas is found underneath the stomach and points toward the spleen.  
(credit: modification of work by NCI)

The endocrine cells of the pancreas form clusters called pancreatic islets or the **islets of Langerhans**, as visible in the micrograph shown in [\[link\]](#). The pancreatic islets contain two primary cell types: **alpha cells**, which produce the hormone glucagon, and **beta cells**, which produce the hormone insulin. These hormones regulate blood glucose levels. As blood glucose levels decline, alpha cells release glucagon to raise the blood glucose levels by increasing rates of glycogen breakdown and glucose release by the liver. When blood glucose levels rise, such as after a meal, beta cells release insulin to lower blood glucose levels by increasing the rate of glucose uptake in most body cells, and by increasing glycogen synthesis in skeletal muscles and the liver. Together, glucagon and insulin regulate blood glucose levels.



The islets of Langerhans are clusters of endocrine cells found in the pancreas; they stain lighter than surrounding cells. (credit: modification of work by Muhammad T. Tabiin, Christopher P. White, Grant Morahan, and Bernard E. Tuch; scale-bar data from Matt Russell)

## **Pineal Gland**

The pineal gland produces melatonin. The rate of melatonin production is affected by the photoperiod. Collaterals from the visual pathways innervate the pineal gland. During the day photoperiod, little melatonin is produced; however, melatonin production increases during the dark photoperiod (night). In some mammals, melatonin has an inhibitory affect on reproductive functions by decreasing production and maturation of sperm, oocytes, and reproductive organs. Melatonin is an effective antioxidant, protecting the CNS from free radicals such as nitric oxide and hydrogen peroxide. Lastly, melatonin is involved in biological rhythms, particularly circadian rhythms such as the sleep-wake cycle and eating habits.

# Gonads

The gonads—the male testes and female ovaries—produce steroid hormones. The testes produce androgens, testosterone being the most prominent, which allow for the development of secondary sex characteristics and the production of sperm cells. The ovaries produce estradiol and progesterone, which cause secondary sex characteristics and prepare the body for childbirth.

Endocrine Glands and their Associated Hormones		
Endocrine Gland	Associated Hormones	Effect
Hypothalamus	releasing and inhibiting hormones	regulate hormone release from pituitary gland; produce oxytocin; produce uterine contractions and milk secretion in females
	antidiuretic hormone (ADH)	water reabsorption from kidneys; vasoconstriction to increase blood pressure
Pituitary (Anterior)	growth hormone (GH)	promotes growth of body tissues, protein synthesis; metabolic functions
	prolactin (PRL)	promotes milk production

Endocrine Glands and their Associated Hormones		
Endocrine Gland	Associated Hormones	Effect
	thyroid stimulating hormone (TSH)	stimulates thyroid hormone release
	adrenocorticotrophic hormone (ACTH)	stimulates hormone release by adrenal cortex, glucocorticoids
	follicle-stimulating hormone (FSH)	stimulates gamete production (both ova and sperm); secretion of estradiol
	luteinizing hormone (LH)	stimulates androgen production by gonads; ovulation, secretion of progesterone
	melanocyte-stimulating hormone (MSH)	stimulates melanocytes of the skin increasing melanin pigment production.
Pituitary (Posterior)	antidiuretic hormone (ADH)	stimulates water reabsorption by kidneys

Endocrine Glands and their Associated Hormones		
Endocrine Gland	Associated Hormones	Effect
	oxytocin	stimulates uterine contractions during childbirth; milk ejection; stimulates ductus deferens and prostate gland contraction during emission
Thyroid	thyroxine, triiodothyronine	stimulate and maintain metabolism; growth and development
	calcitonin	reduces blood $\text{Ca}^{2+}$ levels
Parathyroid	parathyroid hormone (PTH)	increases blood $\text{Ca}^{2+}$ levels
Adrenal (Cortex)	aldosterone	increases blood $\text{Na}^{+}$ levels; increase $\text{K}^{+}$ secretion
	cortisol, corticosterone, cortisone	increase blood glucose levels; anti-inflammatory effects
Adrenal (Medulla)	epinephrine, norepinephrine	stimulate fight-or-flight response; increase blood glucose levels; increase metabolic activities



<b>Endocrine Glands and their Associated Hormones</b>		
<b>Endocrine Gland</b>	<b>Associated Hormones</b>	<b>Effect</b>
Pancreas	insulin	reduces blood glucose levels
	glucagon	increases blood glucose levels
Pineal gland	melatonin	regulates some biological rhythms and protects CNS from free radicals
Testes	androgens	regulate, promote, increase or maintain sperm production; male secondary sexual characteristics
Ovaries	estrogen	promotes uterine lining growth; female secondary sexual characteristics
	progestins	promote and maintain uterine lining growth

## Organs with Secondary Endocrine Functions

There are several organs whose primary functions are non-endocrine but that also possess endocrine functions. These include the heart, kidneys, intestines, thymus, gonads, and adipose tissue.

The heart possesses endocrine cells in the walls of the atria that are specialized cardiac muscle cells. These cells release the hormone **atrial natriuretic peptide (ANP)** in response to increased blood volume. High blood volume causes the cells to be stretched, resulting in hormone release. ANP acts on the kidneys to reduce the reabsorption of  $\text{Na}^+$ , causing  $\text{Na}^+$  and water to be excreted in the urine. ANP also reduces the amounts of renin released by the kidneys and aldosterone released by the adrenal cortex, further preventing the retention of water. In this way, ANP causes a reduction in blood volume and blood pressure, and reduces the concentration of  $\text{Na}^+$  in the blood.

The gastrointestinal tract produces several hormones that aid in digestion. The endocrine cells are located in the mucosa of the GI tract throughout the stomach and small intestine. Some of the hormones produced include gastrin, secretin, and cholecystokinin, which are secreted in the presence of food, and some of which act on other organs such as the pancreas, gallbladder, and liver. They trigger the release of gastric juices, which help to break down and digest food in the GI tract.

While the adrenal glands associated with the kidneys are major **endocrine glands**, the kidneys themselves also possess endocrine function. Renin is released in response to decreased blood volume or pressure and is part of the renin-angiotensin-aldosterone system that leads to the release of aldosterone. Aldosterone then causes the retention of  $\text{Na}^+$  and water, raising blood volume. The kidneys also release calcitriol, which aids in the absorption of  $\text{Ca}^{2+}$  and phosphate ions. **Erythropoietin (EPO)** is a protein hormone that triggers the formation of red blood cells in the bone marrow. EPO is released in response to low oxygen levels. Because red blood cells are oxygen carriers, increased production results in greater oxygen delivery throughout the body. EPO has been used by athletes to improve performance, as greater oxygen delivery to muscle cells allows for greater endurance. Because red blood cells increase the viscosity of blood, artificially high levels of EPO can cause severe health risks.

The **thymus** is found behind the sternum; it is most prominent in infants, becoming smaller in size through adulthood. The thymus produces

hormones referred to as thymosins, which contribute to the development of the immune response.

Adipose tissue is a connective tissue found throughout the body. It produces the hormone **leptin** in response to food intake. Leptin increases the activity of anorexigenic neurons and decreases that of orexigenic neurons, producing a feeling of satiety after eating, thus affecting appetite and reducing the urge for further eating. Leptin is also associated with reproduction. It must be present for GnRH and gonadotropin synthesis to occur. Extremely thin females may enter puberty late; however, if adipose levels increase, more leptin will be produced, improving fertility.

## Section Summary

The pituitary gland is located at the base of the brain and is attached to the hypothalamus by the infundibulum. The anterior pituitary receives products from the hypothalamus by the hypophyseal portal system and produces six hormones. The posterior pituitary is an extension of the brain and releases hormones (antidiuretic hormone and oxytocin) produced by the hypothalamus.

The thyroid gland is located in the neck and is composed of two lobes connected by the isthmus. The thyroid is made up of follicle cells that produce the hormones thyroxine and triiodothyronine. Parafollicular cells of the thyroid produce calcitonin. The parathyroid glands lie on the posterior surface of the thyroid gland and produce parathyroid hormone.

The adrenal glands are located on top of the kidneys and consist of the renal cortex and renal medulla. The adrenal cortex is the outer part of the adrenal gland and produces the corticosteroids, glucocorticoids, and mineralocorticoids. The adrenal medulla is the inner part of the adrenal gland and produces the catecholamines epinephrine and norepinephrine.

The pancreas lies in the abdomen between the stomach and the small intestine. Clusters of endocrine cells in the pancreas form the islets of Langerhans, which are composed of alpha cells that release glucagon and beta cells that release insulin.

Some organs possess endocrine activity as a secondary function but have another primary function. The heart produces the hormone atrial natriuretic peptide, which functions to reduce blood volume, pressure, and  $\text{Na}^+$  concentration. The gastrointestinal tract produces various hormones that aid in digestion. The kidneys produce renin, calcitriol, and erythropoietin. Adipose tissue produces leptin, which promotes satiety signals in the brain.

## Review Questions

### Exercise:

**Problem:** Which endocrine glands are associated with the kidneys?

- a. thyroid glands
- b. pituitary glands
- c. adrenal glands
- d. gonads

---

### Solution:

C

### Exercise:

### Problem:

Which of the following hormones is not produced by the anterior pituitary?

- a. oxytocin
- b. growth hormone
- c. prolactin
- d. thyroid-stimulating hormone

---

### Solution:

A

## Free Response

### Exercise:

**Problem:** What does aldosterone regulate, and how is it stimulated?

---

#### **Solution:**

The main mineralocorticoid is aldosterone, which regulates the concentration of ions in urine, sweat, and saliva. Aldosterone release from the adrenal cortex is stimulated by a decrease in blood concentrations of sodium ions, blood volume, or blood pressure, or an increase in blood potassium levels.

### Exercise:

#### **Problem:**

The adrenal medulla contains two types of secretory cells, what are they and what are their functions?

---

#### **Solution:**

The adrenal medulla contains two types of secretory cells, one that produces epinephrine (adrenaline) and another that produces norepinephrine (noradrenaline). Epinephrine is the primary adrenal medulla hormone accounting for 75–80 percent of its secretions. Epinephrine and norepinephrine increase heart rate, breathing rate, cardiac muscle contractions, and blood glucose levels. They also accelerate the breakdown of glucose in skeletal muscles and stored fats in adipose tissue. The release of epinephrine and norepinephrine is stimulated by neural impulses from the sympathetic nervous system. These neural impulses originate from the hypothalamus in response to stress to prepare the body for the fight-or-flight response.

## **Glossary**

adrenal cortex

outer portion of adrenal glands that produces corticosteroids

adrenal gland

endocrine glands associated with the kidneys

adrenal medulla

inner portion of adrenal glands that produces epinephrine and norepinephrine

alpha cell

endocrine cell of the pancreatic islets that produces the hormone glucagon

anterior pituitary

portion of the pituitary gland that produces six hormones; also called adenohypophysis

atrial natriuretic peptide (ANP)

hormone produced by the heart to reduce blood volume, pressure, and  $\text{Na}^+$  concentration

beta cell

endocrine cell of the pancreatic islets that produces the hormone insulin

colloid

fluid inside the thyroid gland that contains the glycoprotein thyroglobulin

endocrine gland

gland that secretes hormones into the surrounding interstitial fluid, which then diffuse into blood and are carried to various organs and tissues within the body

erythropoietin (EPO)

hormone produced by the kidneys to stimulate red blood cell production in the bone marrow

hypophyseal portal system

system of blood vessels that carries hormones from the hypothalamus to the anterior pituitary

islets of Langerhans (pancreatic islets)

endocrine cells of the pancreas

isthmus

tissue mass that connects the two lobes of the thyroid gland

leptin

hormone produced by adipose tissue that promotes feelings of satiety and reduces hunger

pancreas

organ located between the stomach and the small intestine that contains exocrine and endocrine cells

parafollicular cell

thyroid cell that produces the hormone calcitonin

parathyroid gland

gland located on the surface of the thyroid that produces parathyroid hormone

pituitary gland

endocrine gland located at the base of the brain composed of an anterior and posterior region; also called hypophysis

pituitary stalk

(also, infundibulum) stalk that connects the pituitary gland to the hypothalamus

posterior pituitary

extension of the brain that releases hormones produced by the hypothalamus; along with the infundibulum, it is also referred to as the neurohypophysis

thymus

gland located behind the sternum that produces thymosin hormones that contribute to the development of the immune system

thyroid gland

endocrine gland located in the neck that produces thyroid hormones thyroxine and triiodothyronine



## Types of Hormones

By the end of this section, you will be able to:

- List the different types of hormones
- Explain their role in maintaining homeostasis

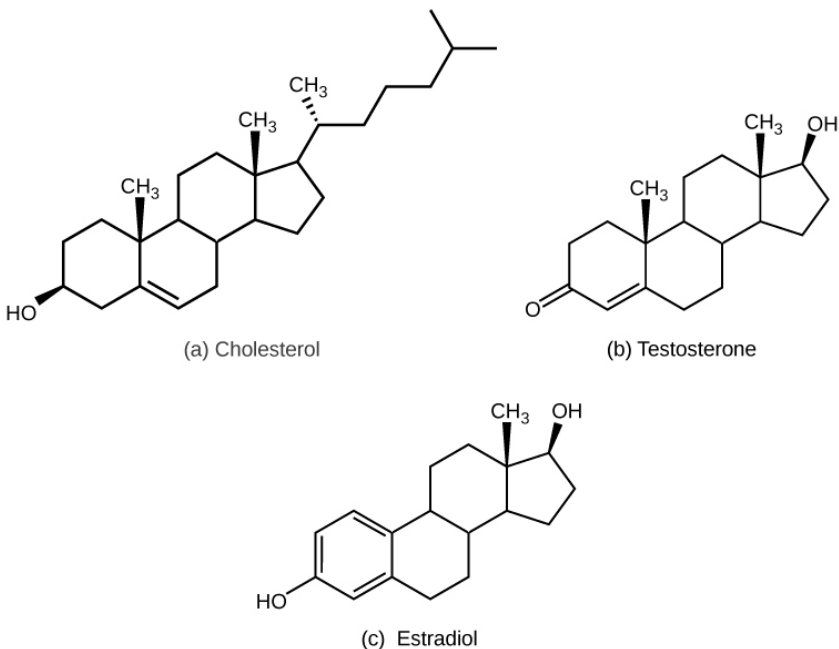
Maintaining homeostasis within the body requires the coordination of many different systems and organs. Communication between neighboring cells, and between cells and tissues in distant parts of the body, occurs through the release of chemicals called hormones. Hormones are released into body fluids (usually blood) that carry these chemicals to their target cells. At the target cells, which are cells that have a receptor for a signal or ligand from a signal cell, the hormones elicit a response. The cells, tissues, and organs that secrete hormones make up the endocrine system. Examples of glands of the endocrine system include the adrenal glands, which produce hormones such as epinephrine and norepinephrine that regulate responses to stress, and the thyroid gland, which produces thyroid hormones that regulate metabolic rates.

Although there are many different hormones in the human body, they can be divided into three classes based on their chemical structure: lipid-derived, amino acid-derived, and peptide (peptide and proteins) hormones. One of the key distinguishing features of lipid-derived hormones is that they can diffuse across plasma membranes whereas the amino acid-derived and peptide hormones cannot.

## Lipid-Derived Hormones (or Lipid-soluble Hormones)

Most **lipid hormones** are derived from cholesterol and thus are structurally similar to it, as illustrated in [\[link\]](#). The primary class of lipid hormones in humans is the steroid hormones. Chemically, these hormones are usually ketones or alcohols; their chemical names will end in “-ol” for alcohols or “-one” for ketones. Examples of steroid hormones include estradiol, which is an **estrogen**, or female sex hormone, and testosterone, which is an androgen, or male sex hormone. These two hormones are released by the female and male reproductive organs, respectively. Other steroid hormones include aldosterone and cortisol, which are released by the adrenal glands

along with some other types of androgens. Steroid hormones are insoluble in water, and they are transported by transport proteins in blood. As a result, they remain in circulation longer than peptide hormones. For example, cortisol has a half-life of 60 to 90 minutes, while epinephrine, an amino acid derived-hormone, has a half-life of approximately one minute.

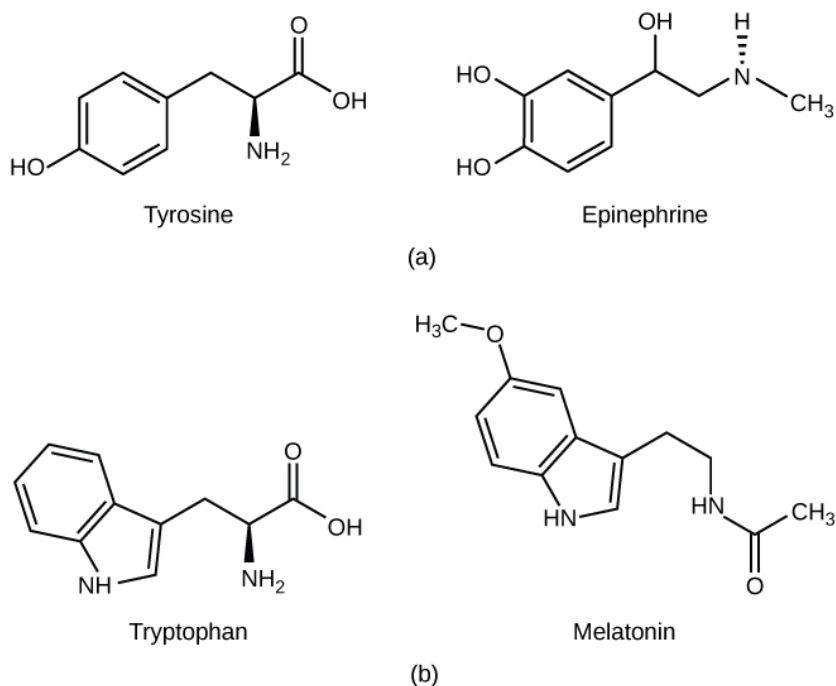


The structures shown here represent (a) cholesterol, plus the steroid hormones (b) testosterone and (c) estradiol.

## Amino Acid-Derived Hormones

The **amino acid-derived hormones** are relatively small molecules that are derived from the amino acids tyrosine and tryptophan, shown in [\[link\]](#). If a hormone is amino acid-derived, its chemical name will end in “-ine”. Examples of amino acid-derived hormones include epinephrine and norepinephrine, which are synthesized in the medulla of the adrenal glands,

and thyroxine, which is produced by the thyroid gland. The pineal gland in the brain makes and secretes melatonin which regulates sleep cycles.



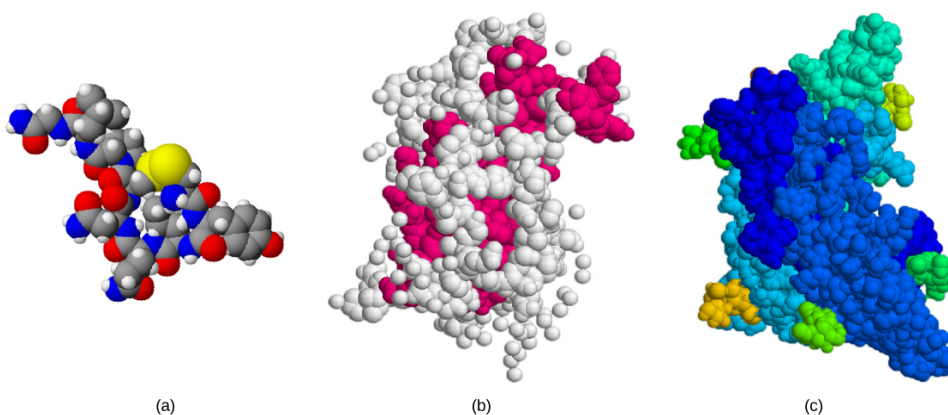
(a) The hormone epinephrine, which triggers the fight-or-flight response, is derived from the amino acid tyrosine. (b) The hormone melatonin, which regulates circadian rhythms, is derived from the amino acid tryptophan.

## Peptide Hormones

The structure of **peptide hormones** is that of a polypeptide chain (chain of amino acids). The peptide hormones include molecules that are short polypeptide chains, such as antidiuretic hormone and oxytocin produced in the brain and released into the blood in the posterior pituitary gland. This class also includes small proteins, like growth hormones produced by the

pituitary, and large glycoproteins such as follicle-stimulating hormone produced by the pituitary. [\[link\]](#) illustrates these peptide hormones.

Secreted peptides like insulin are stored within vesicles in the cells that synthesize them. They are then released in response to stimuli such as high blood glucose levels in the case of insulin. Amino acid-derived and polypeptide hormones are water-soluble and insoluble in lipids. These hormones cannot pass through plasma membranes of cells; therefore, their receptors are found on the surface of the target cells.



The structures of peptide hormones (a) oxytocin, (b) growth hormone, and (c) follicle-stimulating hormone are shown. These peptide hormones are much larger than those derived from cholesterol or amino acids.

**Note:****Career Connection****Endocrinologist**

An endocrinologist is a medical doctor who specializes in treating disorders of the endocrine glands, hormone systems, and glucose and lipid metabolic pathways. An endocrine surgeon specializes in the surgical treatment of endocrine diseases and glands. Some of the diseases that are

managed by endocrinologists: disorders of the pancreas (diabetes mellitus), disorders of the pituitary (gigantism, acromegaly, and pituitary dwarfism), disorders of the thyroid gland (goiter and Graves' disease), and disorders of the adrenal glands (Cushing's disease and Addison's disease).

Endocrinologists are required to assess patients and diagnose endocrine disorders through extensive use of laboratory tests. Many endocrine diseases are diagnosed using tests that stimulate or suppress endocrine organ functioning. Blood samples are then drawn to determine the effect of stimulating or suppressing an endocrine organ on the production of hormones. For example, to diagnose diabetes mellitus, patients are required to fast for 12 to 24 hours. They are then given a sugary drink, which stimulates the pancreas to produce insulin to decrease blood glucose levels. A blood sample is taken one to two hours after the sugar drink is consumed. If the pancreas is functioning properly, the blood glucose level will be within a normal range. Another example is the A1C test, which can be performed during blood screening. The A1C test measures average blood glucose levels over the past two to three months by examining how well the blood glucose is being managed over a long time.

Once a disease has been diagnosed, endocrinologists can prescribe lifestyle changes and/or medications to treat the disease. Some cases of diabetes mellitus can be managed by exercise, weight loss, and a healthy diet; in other cases, medications may be required to enhance insulin release. If the disease cannot be controlled by these means, the endocrinologist may prescribe insulin injections.

In addition to clinical practice, endocrinologists may also be involved in primary research and development activities. For example, ongoing islet transplant research is investigating how healthy pancreas islet cells may be transplanted into diabetic patients. Successful islet transplants may allow patients to stop taking insulin injections.

## Section Summary

There are three basic types of hormones: lipid-derived, amino acid-derived, and peptide. Lipid-derived hormones are structurally similar to cholesterol and include steroid hormones such as estradiol and testosterone. Amino

acid-derived hormones are relatively small molecules and include the adrenal hormones epinephrine and norepinephrine. Peptide hormones are polypeptide chains or proteins and include the pituitary hormones, antidiuretic hormone (vasopressin), and oxytocin.

## Review Questions

### Exercise:

#### Problem:

A newly discovered hormone contains four amino acids linked together. Under which chemical class would this hormone be classified?

- a. lipid-derived hormone
- b. amino acid-derived hormone
- c. peptide hormone
- d. glycoprotein

---

#### Solution:

C

### Exercise:

#### Problem:

Which class of hormones can diffuse through plasma membranes?

- a. lipid-derived hormones
- b. amino acid-derived hormones
- c. peptide hormones
- d. glycoprotein hormones

---

#### Solution:

A

## Free Response

### Exercise:

#### Problem:

Although there are many different hormones in the human body, they can be divided into three classes based on their chemical structure. What are these classes and what is one factor that distinguishes them?

---

#### Solution:

Although there are many different hormones in the human body, they can be divided into three classes based on their chemical structure: lipid-derived, amino acid-derived, and peptide hormones. One of the key distinguishing features of the lipid-derived hormones is that they can diffuse across plasma membranes whereas the amino acid-derived and peptide hormones cannot.

### Exercise:

**Problem:** Where is insulin stored, and why would it be released?

---

#### Solution:

Secreted peptides such as insulin are stored within vesicles in the cells that synthesize them. They are then released in response to stimuli such as high blood glucose levels in the case of insulin.

## Glossary

amino acid-derived hormone  
hormone derived from amino acids

lipid-derived hormone

hormone derived mostly from cholesterol

peptide hormone

hormone composed of a polypeptide chain



## How Hormones Work

By the end of this section, you will be able to:

- Explain how hormones work
- Discuss the role of different types of hormone receptors

Hormones mediate changes in target cells by binding to specific **hormone receptors**. In this way, even though hormones circulate throughout the body and come into contact with many different cell types, they only affect cells that possess the necessary receptors. Receptors for a specific hormone may be found on many different cells or may be limited to a small number of specialized cells. For example, thyroid hormones act on many different tissue types, stimulating metabolic activity throughout the body. Cells can have many receptors for the same hormone but often also possess receptors for different types of hormones. The number of receptors that respond to a hormone determines the cell's sensitivity to that hormone, and the resulting cellular response. Additionally, the number of receptors that respond to a hormone can change over time, resulting in increased or decreased cell sensitivity. In **up-regulation**, the number of receptors increases in response to rising hormone levels, making the cell more sensitive to the hormone and allowing for more cellular activity. When the number of receptors decreases in response to rising hormone levels, called **down-regulation**, cellular activity is reduced.

Receptor binding alters cellular activity and results in an increase or decrease in normal body processes. Depending on the location of the protein receptor on the target cell and the chemical structure of the hormone, hormones can mediate changes directly by binding to **intracellular hormone receptors** and modulating gene transcription, or indirectly by binding to cell surface receptors and stimulating signaling pathways.

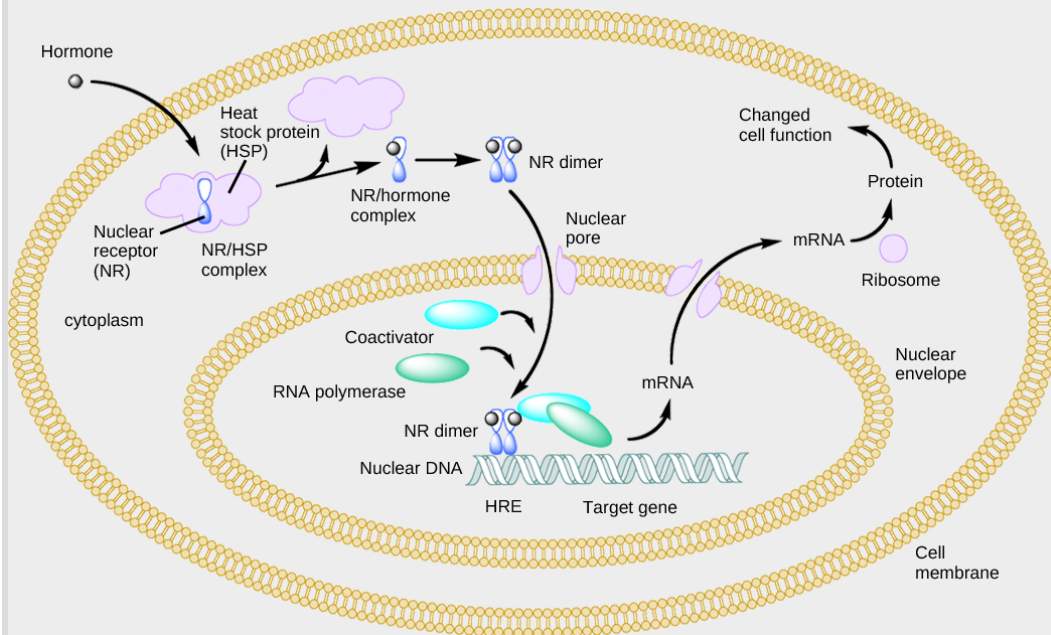
## Intracellular Hormone Receptors

Lipid-derived (soluble) hormones such as steroid hormones diffuse across the membranes of the endocrine cell. Once outside the cell, they bind to transport proteins that keep them soluble in the bloodstream. At the target

cell, the hormones are released from the carrier protein and diffuse across the lipid bilayer of the plasma membrane of cells. The steroid hormones pass through the plasma membrane of a target cell and adhere to intracellular receptors residing in the cytoplasm or in the nucleus. The cell signaling pathways induced by the steroid hormones regulate specific genes on the cell's DNA. The hormones and receptor complex act as transcription regulators by increasing or decreasing the synthesis of mRNA molecules of specific genes. This, in turn, determines the amount of corresponding protein that is synthesized by altering gene expression. This protein can be used either to change the structure of the cell or to produce enzymes that catalyze chemical reactions. In this way, the steroid hormone regulates specific cell processes as illustrated in [\[link\]](#).

### Note:

#### Art Connection



An intracellular nuclear receptor (NR) is located in the cytoplasm bound to a heat shock protein (HSP). Upon hormone binding, the receptor dissociates from the heat shock protein and translocates to the nucleus. In the nucleus, the hormone-receptor complex binds to a DNA sequence called a hormone response element (HRE), which triggers

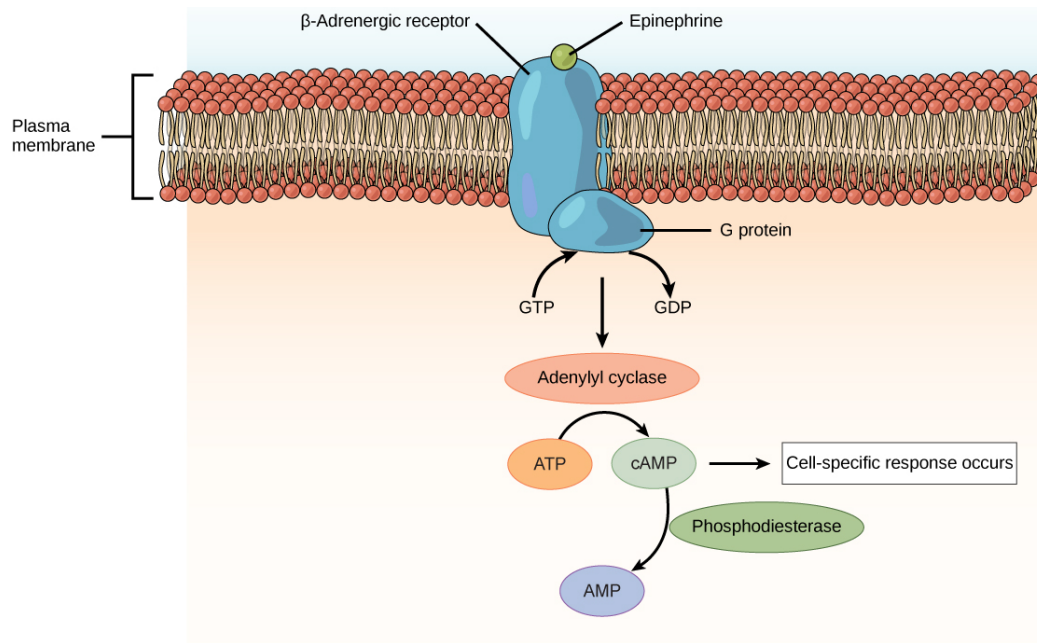
gene transcription and translation. The corresponding protein product can then mediate changes in cell function.

Heat shock proteins (HSP) are so named because they help refold misfolded proteins. In response to increased temperature (a “heat shock”), heat shock proteins are activated by release from the NR/HSP complex. At the same time, transcription of HSP genes is activated. Why do you think the cell responds to a heat shock by increasing the activity of proteins that help refold misfolded proteins?

Other lipid-soluble hormones that are not steroid hormones, such as vitamin D and thyroxine, have receptors located in the nucleus. The hormones diffuse across both the plasma membrane and the nuclear envelope, then bind to receptors in the nucleus. The hormone-receptor complex stimulates transcription of specific genes.

## Plasma Membrane Hormone Receptors

Amino acid derived hormones and polypeptide hormones are not lipid-derived (lipid-soluble) and therefore cannot diffuse through the plasma membrane of cells. Lipid insoluble hormones bind to receptors on the outer surface of the plasma membrane, via **plasma membrane hormone receptors**. Unlike steroid hormones, lipid insoluble hormones do not directly affect the target cell because they cannot enter the cell and act directly on DNA. Binding of these hormones to a cell surface receptor results in activation of a signaling pathway; this triggers intracellular activity and carries out the specific effects associated with the hormone. In this way, nothing passes through the cell membrane; the hormone that binds at the surface remains at the surface of the cell while the intracellular product remains inside the cell. The hormone that initiates the signaling pathway is called a **first messenger**, which activates a second messenger in the cytoplasm, as illustrated in [\[link\]](#).



The amino acid-derived hormones epinephrine and norepinephrine bind to beta-adrenergic receptors on the plasma membrane of cells. Hormone binding to receptor activates a G-protein, which in turn activates adenylyl cyclase, converting ATP to cAMP. cAMP is a second messenger that mediates a cell-specific response. An enzyme called phosphodiesterase breaks down cAMP, terminating the signal.

One very important second messenger is cyclic AMP (cAMP). When a hormone binds to its membrane receptor, a **G-protein** that is associated with the receptor is activated; G-proteins are proteins separate from receptors that are found in the cell membrane. When a hormone is not bound to the receptor, the G-protein is inactive and is bound to guanosine diphosphate, or GDP. When a hormone binds to the receptor, the G-protein is activated by binding guanosine triphosphate, or GTP, in place of GDP. After binding, GTP is hydrolysed by the G-protein into GDP and becomes inactive.

The activated G-protein in turn activates a membrane-bound enzyme called **adenylyl cyclase**. Adenylyl cyclase catalyzes the conversion of ATP to cAMP. cAMP, in turn, activates a group of proteins called protein kinases, which transfer a phosphate group from ATP to a substrate molecule in a process called phosphorylation. The phosphorylation of a substrate molecule changes its structural orientation, thereby activating it. These activated molecules can then mediate changes in cellular processes.

The effect of a hormone is amplified as the signaling pathway progresses. The binding of a hormone at a single receptor causes the activation of many G-proteins, which activates adenylyl cyclase. Each molecule of adenylyl cyclase then triggers the formation of many molecules of cAMP. Further amplification occurs as protein kinases, once activated by cAMP, can catalyze many reactions. In this way, a small amount of hormone can trigger the formation of a large amount of cellular product. To stop hormone activity, cAMP is deactivated by the cytoplasmic enzyme **phosphodiesterase**, or PDE. PDE is always present in the cell and breaks down cAMP to control hormone activity, preventing overproduction of cellular products.

The specific response of a cell to a lipid insoluble hormone depends on the type of receptors that are present on the cell membrane and the substrate molecules present in the cell cytoplasm. Cellular responses to hormone binding of a receptor include altering membrane permeability and metabolic pathways, stimulating synthesis of proteins and enzymes, and activating hormone release.

## Section Summary

Hormones cause cellular changes by binding to receptors on target cells. The number of receptors on a target cell can increase or decrease in response to hormone activity. Hormones can affect cells directly through intracellular hormone receptors or indirectly through plasma membrane hormone receptors.

Lipid-derived (soluble) hormones can enter the cell by diffusing across the plasma membrane and binding to DNA to regulate gene transcription and to

change the cell's activities by inducing production of proteins that affect, in general, the long-term structure and function of the cell. Lipid insoluble hormones bind to receptors on the plasma membrane surface and trigger a signaling pathway to change the cell's activities by inducing production of various cell products that affect the cell in the short-term. The hormone is called a first messenger and the cellular component is called a second messenger. G-proteins activate the second messenger (cyclic AMP), triggering the cellular response. Response to hormone binding is amplified as the signaling pathway progresses. Cellular responses to hormones include the production of proteins and enzymes and altered membrane permeability.

## Art Connections

### Exercise:

#### Problem:

[\[link\]](#) Heat shock proteins (HSP) are so named because they help refold mis-folded proteins. In response to increased temperature (a “heat shock”), heat shock proteins are activated by release from the NR/HSP complex. At the same time, transcription of HSP genes is activated. Why do you think the cell responds to a heat shock by increasing the activity of proteins that help refold misfolded proteins?

---

#### Solution:

[\[link\]](#) Proteins unfold, or denature, at higher temperatures.

## Review Questions

### Exercise:

**Problem:**

A new antagonist molecule has been discovered that binds to and blocks plasma membrane receptors. What effect will this antagonist have on testosterone, a steroid hormone?

- a. It will block testosterone from binding to its receptor.
- b. It will block testosterone from activating cAMP signaling.
- c. It will increase testosterone-mediated signaling.
- d. It will not affect testosterone-mediated signaling.

---

**Solution:**

D

**Exercise:****Problem:**

What effect will a cAMP inhibitor have on a peptide hormone-mediated signaling pathway?

- a. It will prevent the hormone from binding its receptor.
- b. It will prevent activation of a G-protein.
- c. It will prevent activation of adenylate cyclase.
- d. It will prevent activation of protein kinases.

---

**Solution:**

D

**Free Response****Exercise:**

**Problem:** Name two important functions of hormone receptors.

---

**Solution:**

The number of receptors that respond to a hormone can change, resulting in increased or decreased cell sensitivity. The number of receptors can increase in response to rising hormone levels, called up-regulation, making the cell more sensitive to the hormone and allowing for more cellular activity. The number of receptors can also decrease in response to rising hormone levels, called down-regulation, leading to reduced cellular activity.

**Exercise:**

**Problem:**How can hormones mediate changes?

---

**Solution:**

Depending on the location of the protein receptor on the target cell and the chemical structure of the hormone, hormones can mediate changes directly by binding to intracellular receptors and modulating gene transcription, or indirectly by binding to cell surface receptors and stimulating signaling pathways.

**Glossary**

adenylate cyclase

an enzyme that catalyzes the conversion of ATP to cyclic AMP

down-regulation

a decrease in the number of hormone receptors in response to increased hormone levels

first messenger

the hormone that binds to a plasma membrane hormone receptor to trigger a signal transduction pathway

G-protein



a membrane protein activated by the hormone first messenger to activate formation of cyclic AMP

hormone receptor

the cellular protein that binds to a hormone

intracellular hormone receptor

a hormone receptor in the cytoplasm or nucleus of a cell

phosphodiesterase (PDE)

enzyme that deactivates cAMP, stopping hormone activity

plasma membrane hormone receptor

a hormone receptor on the surface of the plasma membrane of a cell

up-regulation

an increase in the number of hormone receptors in response to increased hormone levels

## Regulation of Body Processes

By the end of this section, you will be able to:

- Explain how hormones regulate the excretory system
- Discuss the role of hormones in the reproductive system
- Describe how hormones regulate metabolism
- Explain the role of hormones in different diseases

Hormones have a wide range of effects and modulate many different body processes. The key regulatory processes that will be examined here are those affecting the excretory system, the reproductive system, metabolism, blood calcium concentrations, growth, and the stress response.

### Hormonal Regulation of the Excretory System

Maintaining a proper water balance in the body is important to avoid dehydration or over-hydration (hyponatremia). The water concentration of the body is monitored by **osmoreceptors** in the hypothalamus, which detect the concentration of electrolytes in the extracellular fluid. The concentration of electrolytes in the blood rises when there is water loss caused by excessive perspiration, inadequate water intake, or low blood volume due to blood loss. An increase in blood electrolyte levels results in a neuronal signal being sent from the osmoreceptors in hypothalamic nuclei. The pituitary gland has two components: anterior and posterior. The anterior pituitary is composed of glandular cells that secrete protein hormones. The posterior pituitary is an extension of the hypothalamus. It is composed largely of neurons that are continuous with the hypothalamus.

The hypothalamus produces a polypeptide hormone known as **antidiuretic hormone (ADH)**, which is transported to and released from the posterior pituitary gland. The principal action of ADH is to regulate the amount of water excreted by the kidneys. As ADH (which is also known as vasopressin) causes direct water reabsorption from the kidney tubules, salts and wastes are concentrated in what will eventually be excreted as urine. The hypothalamus controls the mechanisms of ADH secretion, either by regulating blood volume or the concentration of water in the blood. Dehydration or physiological stress can cause an increase of osmolarity

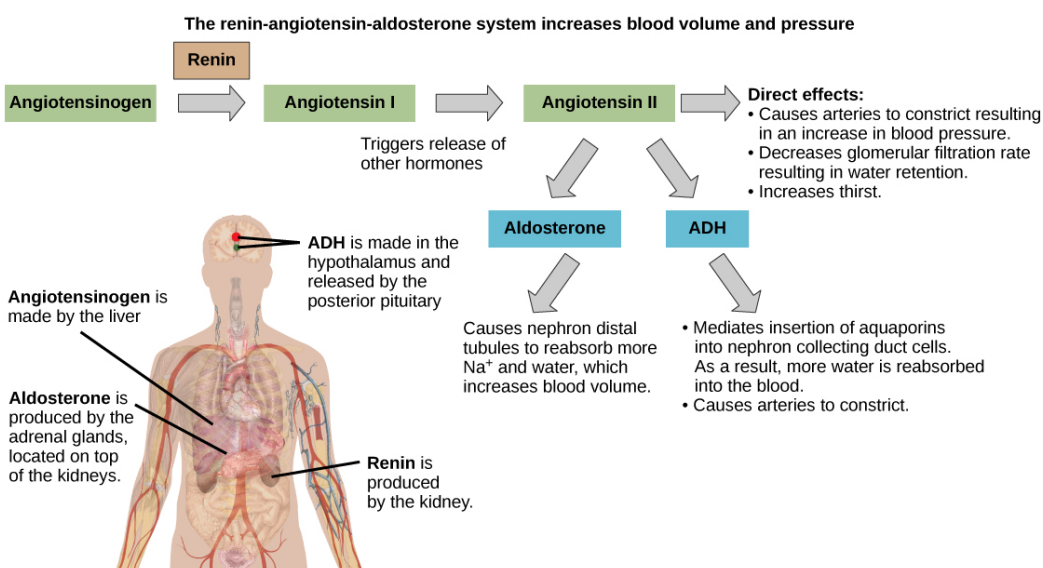
above 300 mOsm/L, which in turn, raises ADH secretion and water will be retained, causing an increase in blood pressure. ADH travels in the bloodstream to the kidneys. Once at the kidneys, ADH changes the kidneys to become more permeable to water by temporarily inserting water channels, aquaporins, into the kidney tubules. Water moves out of the kidney tubules through the aquaporins, reducing urine volume. The water is reabsorbed into the capillaries lowering blood osmolarity back toward normal. As blood osmolarity decreases, a negative feedback mechanism reduces osmoreceptor activity in the hypothalamus, and ADH secretion is reduced. ADH release can be reduced by certain substances, including alcohol, which can cause increased urine production and dehydration.

Chronic underproduction of ADH or a mutation in the ADH receptor results in **diabetes insipidus**. If the posterior pituitary does not release enough ADH, water cannot be retained by the kidneys and is lost as urine. This causes increased thirst, but water taken in is lost again and must be continually consumed. If the condition is not severe, dehydration may not occur, but severe cases can lead to electrolyte imbalances due to dehydration.

Another hormone responsible for maintaining electrolyte concentrations in extracellular fluids is **aldosterone**, a steroid hormone that is produced by the adrenal cortex. In contrast to ADH, which promotes the reabsorption of water to maintain proper water balance, aldosterone maintains proper water balance by enhancing  $\text{Na}^+$  reabsorption and  $\text{K}^+$  secretion from extracellular fluid of the cells in kidney tubules. Because it is produced in the cortex of the adrenal gland and affects the concentrations of minerals  $\text{Na}^+$  and  $\text{K}^+$ , aldosterone is referred to as a **mineralocorticoid**, a corticosteroid that affects ion and water balance. Aldosterone release is stimulated by a decrease in blood sodium levels, blood volume, or blood pressure, or an increase in blood potassium levels. It also prevents the loss of  $\text{Na}^+$  from sweat, saliva, and gastric juice. The reabsorption of  $\text{Na}^+$  also results in the osmotic reabsorption of water, which alters blood volume and blood pressure.

Aldosterone production can be stimulated by low blood pressure, which triggers a sequence of chemical release, as illustrated in [\[link\]](#). When blood

pressure drops, the renin-angiotensin-aldosterone system (RAAS) is activated. Cells in the juxtaglomerular apparatus, which regulates the functions of the nephrons of the kidney, detect this and release **renin**. Renin, an enzyme, circulates in the blood and reacts with a plasma protein produced by the liver called angiotensinogen. When angiotensinogen is cleaved by renin, it produces angiotensin I, which is then converted into angiotensin II in the lungs. Angiotensin II functions as a hormone and then causes the release of the hormone aldosterone by the adrenal cortex, resulting in increased  $\text{Na}^+$  reabsorption, water retention, and an increase in blood pressure. Angiotensin II in addition to being a potent vasoconstrictor also causes an increase in ADH and increased thirst, both of which help to raise blood pressure.



ADH and aldosterone increase blood pressure and volume.

Angiotensin II stimulates release of these hormones.

Angiotensin II, in turn, is formed when renin cleaves angiotensinogen. (credit: modification of work by Mikael Häggström)

## Hormonal Regulation of the Reproductive System

Regulation of the reproductive system is a process that requires the action of hormones from the pituitary gland, the adrenal cortex, and the gonads. During puberty in both males and females, the hypothalamus produces gonadotropin-releasing hormone (GnRH), which stimulates the production and release of **follicle-stimulating hormone (FSH)** and luteinizing hormone (LH) from the anterior pituitary gland. These hormones regulate the gonads (testes in males and ovaries in females) and therefore are called **gonadotropins**. In both males and females, FSH stimulates gamete production and LH stimulates production of hormones by the gonads. An increase in gonad hormone levels inhibits GnRH production through a negative feedback loop.

## **Regulation of the Male Reproductive System**

In males, FSH stimulates the maturation of sperm cells. FSH production is inhibited by the hormone inhibin, which is released by the testes. LH stimulates production of the sex hormones (**androgens**) by the interstitial cells of the testes and therefore is also called interstitial cell-stimulating hormone.

The most widely known androgen in males is testosterone. Testosterone promotes the production of sperm and masculine characteristics. The adrenal cortex also produces small amounts of testosterone precursor, although the role of this additional hormone production is not fully understood.

### **Note:**

Everyday Connection

**The Dangers of Synthetic Hormones**



Professional baseball player Jason Giambi publically admitted to, and apologized for, his use of anabolic steroids supplied by a trainer. (credit: Bryce Edwards)

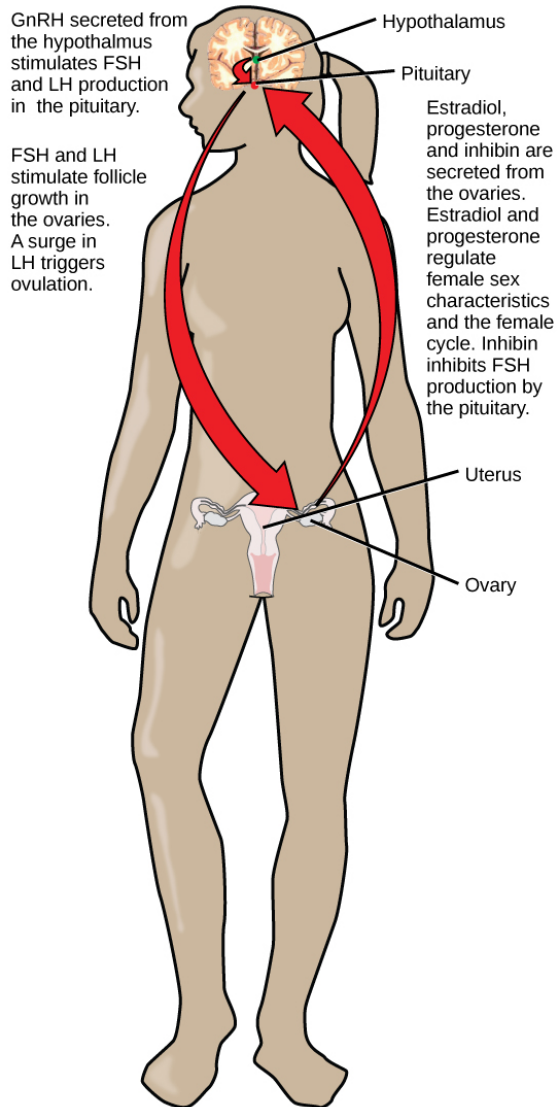
Some athletes attempt to boost their performance by using artificial hormones that enhance muscle performance. Anabolic steroids, a form of the male sex hormone testosterone, are one of the most widely known performance-enhancing drugs. Steroids are used to help build muscle mass. Other hormones that are used to enhance athletic performance include erythropoietin, which triggers the production of red blood cells, and human growth hormone, which can help in building muscle mass. Most performance enhancing drugs are illegal for non-medical purposes. They are also banned by national and international governing bodies including the International Olympic Committee, the U.S. Olympic Committee, the National Collegiate Athletic Association, the Major League Baseball, and the National Football League.

The side effects of synthetic hormones are often significant and non-reversible, and in some cases, fatal. Androgens produce several complications such as liver dysfunctions and liver tumors, prostate gland enlargement, difficulty urinating, premature closure of epiphyseal cartilages, testicular atrophy, infertility, and immune system depression. The physiological strain caused by these substances is often greater than what the body can handle, leading to unpredictable and dangerous effects

and linking their use to heart attacks, strokes, and impaired cardiac function.

## **Regulation of the Female Reproductive System**

In females, FSH stimulates development of egg cells, called ova, which develop in structures called follicles. Follicle cells produce the hormone inhibin, which inhibits FSH production. LH also plays a role in the development of ova, induction of ovulation, and stimulation of estradiol and progesterone production by the ovaries, as illustrated in [\[link\]](#). Estradiol and progesterone are steroid hormones that prepare the body for pregnancy. Estradiol produces secondary sex characteristics in females, while both estradiol and progesterone regulate the menstrual cycle.



Hormonal regulation of the female reproductive system involves hormones from the hypothalamus, pituitary, and ovaries.

In addition to producing FSH and LH, the anterior portion of the pituitary gland also produces the hormone **prolactin (PRL)** in females. Prolactin stimulates the production of milk by the mammary glands following childbirth. Prolactin levels are regulated by the hypothalamic hormones



**prolactin-releasing hormone (PRH)** and **prolactin-inhibiting hormone (PIH)**, which is now known to be dopamine. PRH stimulates the release of prolactin and PIH inhibits it.

The posterior pituitary releases the hormone **oxytocin**, which stimulates uterine contractions during childbirth. The uterine smooth muscles are not very sensitive to oxytocin until late in pregnancy when the number of oxytocin receptors in the uterus peaks. Stretching of tissues in the uterus and cervix stimulates oxytocin release during childbirth. Contractions increase in intensity as blood levels of oxytocin rise via a positive feedback mechanism until the birth is complete. Oxytocin also stimulates the contraction of myoepithelial cells around the milk-producing mammary glands. As these cells contract, milk is forced from the secretory alveoli into milk ducts and is ejected from the breasts in milk ejection (“let-down”) reflex. Oxytocin release is stimulated by the suckling of an infant, which triggers the synthesis of oxytocin in the hypothalamus and its release into circulation at the posterior pituitary.

## **Hormonal Regulation of Metabolism**

Blood glucose levels vary widely over the course of a day as periods of food consumption alternate with periods of fasting. Insulin and glucagon are the two hormones primarily responsible for maintaining homeostasis of blood glucose levels. Additional regulation is mediated by the thyroid hormones.

### **Regulation of Blood Glucose Levels by Insulin and Glucagon**

Cells of the body require nutrients in order to function, and these nutrients are obtained through feeding. In order to manage nutrient intake, storing excess intake and utilizing reserves when necessary, the body uses hormones to moderate energy stores. **Insulin** is produced by the beta cells of the pancreas, which are stimulated to release insulin as blood glucose levels rise (for example, after a meal is consumed). Insulin lowers blood glucose levels by enhancing the rate of glucose uptake and utilization by

target cells, which use glucose for ATP production. It also stimulates the liver to convert glucose to glycogen, which is then stored by cells for later use. Insulin also increases glucose transport into certain cells, such as muscle cells and the liver. This results from an insulin-mediated increase in the number of glucose transporter proteins in cell membranes, which remove glucose from circulation by facilitated diffusion. As insulin binds to its target cell via insulin receptors and signal transduction, it triggers the cell to incorporate glucose transport proteins into its membrane. This allows glucose to enter the cell, where it can be used as an energy source. However, this does not occur in all cells: some cells, including those in the kidneys and brain, can access glucose without the use of insulin. Insulin also stimulates the conversion of glucose to fat in adipocytes and the synthesis of proteins. These actions mediated by insulin cause blood glucose concentrations to fall, called a hypoglycemic “low sugar” effect, which inhibits further insulin release from beta cells through a negative feedback loop.

**Note:**

Link to Learning

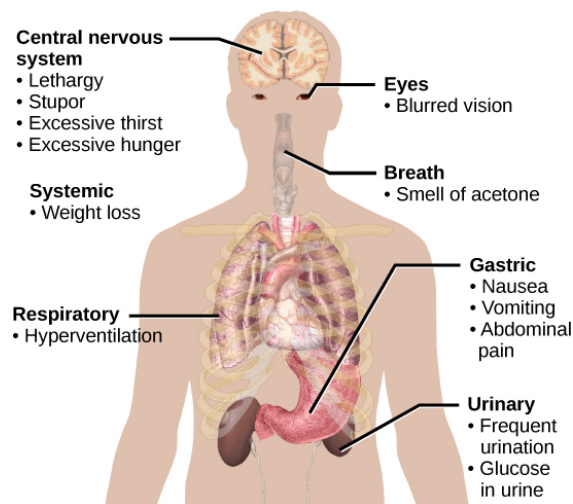


This animation describe the role of insulin and the pancreas in diabetes.

<https://www.openstaxcollege.org/l/insulin>

Impaired insulin function can lead to a condition called **diabetes mellitus**, the main symptoms of which are illustrated in [\[link\]](#). This can be caused by low levels of insulin production by the beta cells of the pancreas, or by reduced sensitivity of tissue cells to insulin. This prevents glucose from being absorbed by cells, causing high levels of blood glucose, or

**hyperglycemia** (high sugar). High blood glucose levels make it difficult for the kidneys to recover all the glucose from nascent urine, resulting in glucose being lost in urine. High glucose levels also result in less water being reabsorbed by the kidneys, causing high amounts of urine to be produced; this may result in dehydration. Over time, high blood glucose levels can cause nerve damage to the eyes and peripheral body tissues, as well as damage to the kidneys and cardiovascular system. Oversecretion of insulin can cause **hypoglycemia**, low blood glucose levels. This causes insufficient glucose availability to cells, often leading to muscle weakness, and can sometimes cause unconsciousness or death if left untreated.



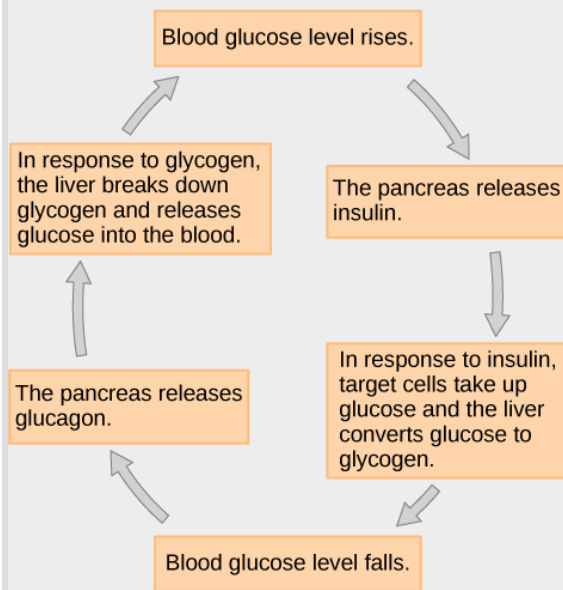
The main symptoms of diabetes are shown. (credit: modification of work by Mikael Häggström)

When blood glucose levels decline below normal levels, for example between meals or when glucose is utilized rapidly during exercise, the hormone **glucagon** is released from the alpha cells of the pancreas. Glucagon raises blood glucose levels, eliciting what is called a hyperglycemic effect, by stimulating the breakdown of glycogen to glucose

in skeletal muscle cells and liver cells in a process called **glycogenolysis**. Glucose can then be utilized as energy by muscle cells and released into circulation by the liver cells. Glucagon also stimulates absorption of amino acids from the blood by the liver, which then converts them to glucose. This process of glucose synthesis is called **gluconeogenesis**. Glucagon also stimulates adipose cells to release fatty acids into the blood. These actions mediated by glucagon result in an increase in blood glucose levels to normal homeostatic levels. Rising blood glucose levels inhibit further glucagon release by the pancreas via a negative feedback mechanism. In this way, insulin and glucagon work together to maintain homeostatic glucose levels, as shown in [\[link\]](#).

**Note:**

**Art Connection**



Insulin and glucagon regulate blood glucose levels.

Pancreatic tumors may cause excess secretion of glucagon. Type I diabetes results from the failure of the pancreas to produce insulin. Which of the following statement about these two conditions is true?

- a. A pancreatic tumor and type I diabetes will have the opposite effects on blood sugar levels.
- b. A pancreatic tumor and type I diabetes will both cause hyperglycemia.
- c. A pancreatic tumor and type I diabetes will both cause hypoglycemia.
- d. Both pancreatic tumors and type I diabetes result in the inability of cells to take up glucose.

## Regulation of Blood Glucose Levels by Thyroid Hormones

The basal metabolic rate, which is the amount of calories required by the body at rest, is determined by two hormones produced by the thyroid gland: **thyroxine**, also known as tetraiodothyronine or  $T_4$ , and **triiodothyronine**, also known as  $T_3$ . These hormones affect nearly every cell in the body except for the adult brain, uterus, testes, blood cells, and spleen. They are transported across the plasma membrane of target cells and bind to receptors on the mitochondria resulting in increased ATP production. In the nucleus,  $T_3$  and  $T_4$  activate genes involved in energy production and glucose oxidation. This results in increased rates of metabolism and body heat production, which is known as the hormone's calorogenic effect.

$T_3$  and  $T_4$  release from the thyroid gland is stimulated by **thyroid-stimulating hormone (TSH)**, which is produced by the anterior pituitary. TSH binding at the receptors of the follicle of the thyroid triggers the production of  $T_3$  and  $T_4$  from a glycoprotein called **thyroglobulin**. Thyroglobulin is present in the follicles of the thyroid, and is converted into thyroid hormones with the addition of iodine. Iodine is formed from iodide ions that are actively transported into the thyroid follicle from the bloodstream. A peroxidase enzyme then attaches the iodine to the tyrosine amino acid found in thyroglobulin.  $T_3$  has three iodine ions attached, while  $T_4$  has four iodine ions attached.  $T_3$  and  $T_4$  are then released into the bloodstream, with  $T_4$  being released in much greater amounts than  $T_3$ . As  $T_3$  is more active than  $T_4$  and is responsible for most of the effects of thyroid hormones, tissues of the body convert  $T_4$  to  $T_3$  by the removal of an iodine ion. Most of the released  $T_3$  and  $T_4$  becomes attached to transport proteins in the bloodstream and is unable to cross the plasma membrane of

cells. These protein-bound molecules are only released when blood levels of the unattached hormone begin to decline. In this way, a week's worth of reserve hormone is maintained in the blood. Increased  $T_3$  and  $T_4$  levels in the blood inhibit the release of TSH, which results in lower  $T_3$  and  $T_4$  release from the thyroid.

The follicular cells of the thyroid require iodides (anions of iodine) in order to synthesize  $T_3$  and  $T_4$ . Iodides obtained from the diet are actively transported into follicle cells resulting in a concentration that is approximately 30 times higher than in blood. The typical diet in North America provides more iodine than required due to the addition of iodide to table salt. Inadequate iodine intake, which occurs in many developing countries, results in an inability to synthesize  $T_3$  and  $T_4$  hormones. The thyroid gland enlarges in a condition called **goiter**, which is caused by overproduction of TSH without the formation of thyroid hormone. Thyroglobulin is contained in a fluid called colloid, and TSH stimulation results in higher levels of colloid accumulation in the thyroid. In the absence of iodine, this is not converted to thyroid hormone, and colloid begins to accumulate more and more in the thyroid gland, leading to goiter.

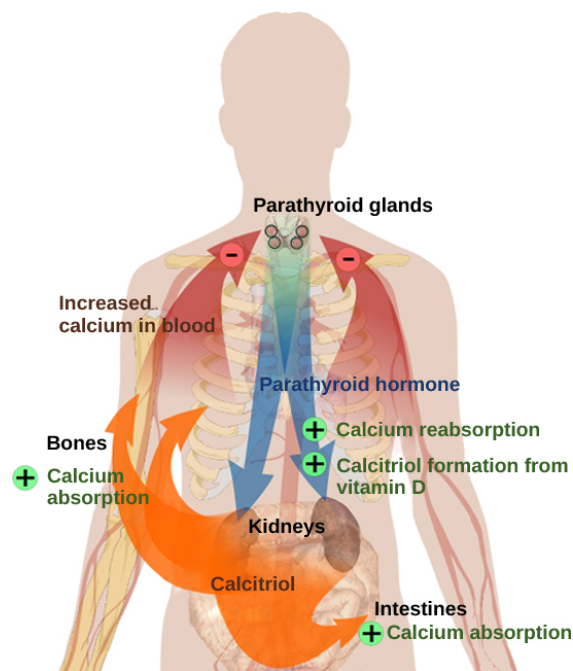
Disorders can arise from both the underproduction and overproduction of thyroid hormones. **Hypothyroidism**, underproduction of the thyroid hormones, can cause a low metabolic rate leading to weight gain, sensitivity to cold, and reduced mental activity, among other symptoms. In children, hypothyroidism can cause cretinism, which can lead to mental retardation and growth defects. **Hyperthyroidism**, the overproduction of thyroid hormones, can lead to an increased metabolic rate and its effects: weight loss, excess heat production, sweating, and an increased heart rate. Graves' disease is one example of a hyperthyroid condition.

## **Hormonal Control of Blood Calcium Levels**

Regulation of blood calcium concentrations is important for generation of muscle contractions and nerve impulses, which are electrically stimulated. If calcium levels get too high, membrane permeability to sodium decreases and membranes become less responsive. If calcium levels get too low,

membrane permeability to sodium increases and convulsions or muscle spasms can result.

Blood calcium levels are regulated by **parathyroid hormone (PTH)**, which is produced by the parathyroid glands, as illustrated in [\[link\]](#). PTH is released in response to low blood  $\text{Ca}^{2+}$  levels. PTH increases  $\text{Ca}^{2+}$  levels by targeting the skeleton, the kidneys, and the intestine. In the skeleton, PTH stimulates osteoclasts, which causes bone to be reabsorbed, releasing  $\text{Ca}^{2+}$  from bone into the blood. PTH also inhibits osteoblasts, reducing  $\text{Ca}^{2+}$  deposition in bone. In the intestines, PTH increases dietary  $\text{Ca}^{2+}$  absorption, and in the kidneys, PTH stimulates reabsorption of the  $\text{Ca}^{2+}$ . While PTH acts directly on the kidneys to increase  $\text{Ca}^{2+}$  reabsorption, its effects on the intestine are indirect. PTH triggers the formation of calcitriol, an active form of vitamin D, which acts on the intestines to increase absorption of dietary calcium. PTH release is inhibited by rising blood calcium levels.



Parathyroid hormone (PTH)  
is released in response to low  
blood calcium levels. It  
increases blood calcium

levels by targeting the skeleton, the kidneys, and the intestine. (credit: modification of work by Mikael Häggström)

Hyperparathyroidism results from an overproduction of parathyroid hormone. This results in excessive calcium being removed from bones and introduced into blood circulation, producing structural weakness of the bones, which can lead to deformation and fractures, plus nervous system impairment due to high blood calcium levels. Hypoparathyroidism, the underproduction of PTH, results in extremely low levels of blood calcium, which causes impaired muscle function and may result in tetany (severe sustained muscle contraction).

The hormone **calcitonin**, which is produced by the parafollicular or C cells of the thyroid, has the opposite effect on blood calcium levels as does PTH. Calcitonin decreases blood calcium levels by inhibiting osteoclasts, stimulating osteoblasts, and stimulating calcium excretion by the kidneys. This results in calcium being added to the bones to promote structural integrity. Calcitonin is most important in children (when it stimulates bone growth), during pregnancy (when it reduces maternal bone loss), and during prolonged starvation (because it reduces bone mass loss). In healthy nonpregnant, unstarved adults, the role of calcitonin is unclear.

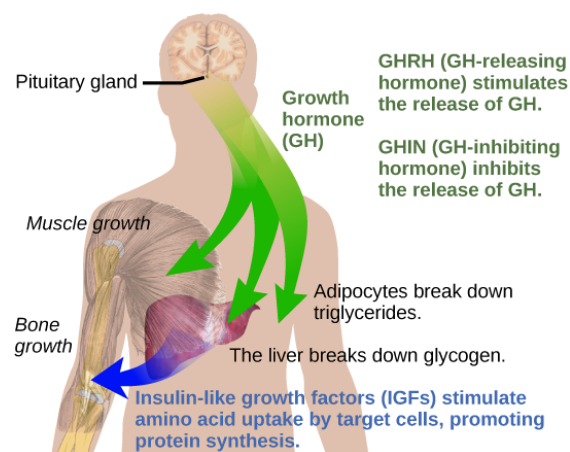
## **Hormonal Regulation of Growth**

Hormonal regulation is required for the growth and replication of most cells in the body. **Growth hormone (GH)**, produced by the anterior portion of the pituitary gland, accelerates the rate of protein synthesis, particularly in skeletal muscle and bones. Growth hormone has direct and indirect mechanisms of action. The first direct action of GH is stimulation of triglyceride breakdown (lipolysis) and release into the blood by adipocytes. This results in a switch by most tissues from utilizing glucose as an energy source to utilizing fatty acids. This process is called a **glucose-sparing**



**effect.** In another direct mechanism, GH stimulates glycogen breakdown in the liver; the glycogen is then released into the blood as glucose. Blood glucose levels increase as most tissues are utilizing fatty acids instead of glucose for their energy needs. The GH mediated increase in blood glucose levels is called a **diabetogenic effect** because it is similar to the high blood glucose levels seen in diabetes mellitus.

The indirect mechanism of GH action is mediated by **insulin-like growth factors (IGFs)** or somatomedins, which are a family of growth-promoting proteins produced by the liver, which stimulates tissue growth. IGFs stimulate the uptake of amino acids from the blood, allowing the formation of new proteins, particularly in skeletal muscle cells, cartilage cells, and other target cells, as shown in [\[link\]](#). This is especially important after a meal, when glucose and amino acid concentration levels are high in the blood. GH levels are regulated by two hormones produced by the hypothalamus. GH release is stimulated by **growth hormone-releasing hormone (GHRH)** and is inhibited by **growth hormone-inhibiting hormone (GHIH)**, also called somatostatin.



Growth hormone directly accelerates the rate of protein synthesis in skeletal muscle and bones. Insulin-like growth factor 1 (IGF-1) is activated by growth hormone

and also allows formation of new proteins in muscle cells and bone. (credit: modification of work by Mikael Häggström)

A balanced production of growth hormone is critical for proper development. Underproduction of GH in adults does not appear to cause any abnormalities, but in children it can result in **pituitary dwarfism**, in which growth is reduced. Pituitary dwarfism is characterized by symmetric body formation. In some cases, individuals are under 30 inches in height. Oversecretion of growth hormone can lead to **gigantism** in children, causing excessive growth. In some documented cases, individuals can reach heights of over eight feet. In adults, excessive GH can lead to **acromegaly**, a condition in which there is enlargement of bones in the face, hands, and feet that are still capable of growth.

## Hormonal Regulation of Stress

When a threat or danger is perceived, the body responds by releasing hormones that will ready it for the “fight-or-flight” response. The effects of this response are familiar to anyone who has been in a stressful situation: increased heart rate, dry mouth, and hair standing up.

### Note:

#### Evolution Connection

#### Fight-or-Flight Response

Interactions of the endocrine hormones have evolved to ensure the body’s internal environment remains stable. Stressors are stimuli that disrupt homeostasis. The sympathetic division of the vertebrate autonomic nervous system has evolved the fight-or-flight response to counter stress-induced disruptions of homeostasis. In the initial alarm phase, the sympathetic nervous system stimulates an increase in energy levels through increased

blood glucose levels. This prepares the body for physical activity that may be required to respond to stress: to either fight for survival or to flee from danger.

However, some stresses, such as illness or injury, can last for a long time. Glycogen reserves, which provide energy in the short-term response to stress, are exhausted after several hours and cannot meet long-term energy needs. If glycogen reserves were the only energy source available, neural functioning could not be maintained once the reserves became depleted due to the nervous system's high requirement for glucose. In this situation, the body has evolved a response to counter long-term stress through the actions of the glucocorticoids, which ensure that long-term energy requirements can be met. The glucocorticoids mobilize lipid and protein reserves, stimulate gluconeogenesis, conserve glucose for use by neural tissue, and stimulate the conservation of salts and water. The mechanisms to maintain homeostasis that are described here are those observed in the human body. However, the fight-or-flight response exists in some form in all vertebrates.

The sympathetic nervous system regulates the stress response via the hypothalamus. Stressful stimuli cause the hypothalamus to signal the adrenal medulla (which mediates short-term stress responses) via nerve impulses, and the adrenal cortex, which mediates long-term stress responses, via the hormone **adrenocorticotrophic hormone (ACTH)**, which is produced by the anterior pituitary.

### **Short-term Stress Response**

When presented with a stressful situation, the body responds by calling for the release of hormones that provide a burst of energy. The hormones **epinephrine** (also known as adrenaline) and **norepinephrine** (also known as noradrenaline) are released by the adrenal medulla. How do these hormones provide a burst of energy? Epinephrine and norepinephrine increase blood glucose levels by stimulating the liver and skeletal muscles to break down glycogen and by stimulating glucose release by liver cells.

Additionally, these hormones increase oxygen availability to cells by increasing the heart rate and dilating the bronchioles. The hormones also prioritize body function by increasing blood supply to essential organs such as the heart, brain, and skeletal muscles, while restricting blood flow to organs not in immediate need, such as the skin, digestive system, and kidneys. Epinephrine and norepinephrine are collectively called catecholamines.

**Note:**

Link to Learning



Watch this [Discovery Channel animation](#) describing the flight-or-flight response.

## Long-term Stress Response

Long-term stress response differs from short-term stress response. The body cannot sustain the bursts of energy mediated by epinephrine and norepinephrine for long times. Instead, other hormones come into play. In a long-term stress response, the hypothalamus triggers the release of ACTH from the anterior pituitary gland. The adrenal cortex is stimulated by ACTH to release steroid hormones called **corticosteroids**. Corticosteroids turn on transcription of certain genes in the nuclei of target cells. They change enzyme concentrations in the cytoplasm and affect cellular metabolism. There are two main corticosteroids: glucocorticoids such as **cortisol**, and mineralocorticoids such as aldosterone. These hormones target the breakdown of fat into fatty acids in the adipose tissue. The fatty acids are released into the bloodstream for other tissues to use for ATP production.

The **glucocorticoids** primarily affect glucose metabolism by stimulating glucose synthesis. Glucocorticoids also have anti-inflammatory properties through inhibition of the immune system. For example, cortisone is used as an anti-inflammatory medication; however, it cannot be used long term as it increases susceptibility to disease due to its immune-suppressing effects.

Mineralocorticoids function to regulate ion and water balance of the body. The hormone aldosterone stimulates the reabsorption of water and sodium ions in the kidney, which results in increased blood pressure and volume.

Hypersecretion of glucocorticoids can cause a condition known as **Cushing's disease**, characterized by a shifting of fat storage areas of the body. This can cause the accumulation of adipose tissue in the face and neck, and excessive glucose in the blood. Hyposecretion of the corticosteroids can cause **Addison's disease**, which may result in bronzing of the skin, hypoglycemia, and low electrolyte levels in the blood.

## Section Summary

Water levels in the body are controlled by antidiuretic hormone (ADH), which is produced in the hypothalamus and triggers the reabsorption of water by the kidneys. Underproduction of ADH can cause diabetes insipidus. Aldosterone, a hormone produced by the adrenal cortex of the kidneys, enhances  $\text{Na}^+$  reabsorption from the extracellular fluids and subsequent water reabsorption by diffusion. The renin-angiotensin-aldosterone system is one way that aldosterone release is controlled.

The reproductive system is controlled by the gonadotropins follicle-stimulating hormone (FSH) and luteinizing hormone (LH), which are produced by the pituitary gland. Gonadotropin release is controlled by the hypothalamic hormone gonadotropin-releasing hormone (GnRH). FSH stimulates the maturation of sperm cells in males and is inhibited by the hormone inhibin, while LH stimulates the production of the androgen testosterone. FSH stimulates egg maturation in females, while LH stimulates the production of estrogens and progesterone. **Estrogens** are a group of steroid hormones produced by the ovaries that trigger the development of secondary sex characteristics in females as well as control

the maturation of the ova. In females, the pituitary also produces prolactin, which stimulates milk production after childbirth, and oxytocin, which stimulates uterine contraction during childbirth and milk let-down during suckling.

Insulin is produced by the pancreas in response to rising blood glucose levels and allows cells to utilize blood glucose and store excess glucose for later use. Diabetes mellitus is caused by reduced insulin activity and causes high blood glucose levels, or hyperglycemia. Glucagon is released by the pancreas in response to low blood glucose levels and stimulates the breakdown of glycogen into glucose, which can be used by the body. The body's basal metabolic rate is controlled by the thyroid hormones thyroxine ( $T_4$ ) and triiodothyronine ( $T_3$ ). The anterior pituitary produces thyroid stimulating hormone (TSH), which controls the release of  $T_3$  and  $T_4$  from the thyroid gland. Iodine is necessary in the production of thyroid hormone, and the lack of iodine can lead to a condition called goiter.

Parathyroid hormone (PTH) is produced by the parathyroid glands in response to low blood  $Ca^{2+}$  levels. The parafollicular cells of the thyroid produce calcitonin, which reduces blood  $Ca^{2+}$  levels. Growth hormone (GH) is produced by the anterior pituitary and controls the growth rate of muscle and bone. GH action is indirectly mediated by insulin-like growth factors (IGFs). Short-term stress causes the hypothalamus to trigger the adrenal medulla to release epinephrine and norepinephrine, which trigger the fight or flight response. Long-term stress causes the hypothalamus to trigger the anterior pituitary to release adrenocorticotrophic hormone (ACTH), which causes the release of corticosteroids, glucocorticoids, and mineralocorticoids, from the adrenal cortex.

## Art Connections

### Exercise:

#### Problem:

[\[link\]](#) Pancreatic tumors may cause excess secretion of glucagon. Type I diabetes results from the failure of the pancreas to produce insulin. Which of the following statement about these two conditions is true?

- a. A pancreatic tumor and type I diabetes will have the opposite effects on blood sugar levels.
- b. A pancreatic tumor and type I diabetes will both cause hyperglycemia.
- c. A pancreatic tumor and type I diabetes will both cause hypoglycemia.
- d. Both pancreatic tumors and type I diabetes result in the inability of cells to take up glucose.

---

**Solution:**

[\[link\]](#) B

## Review Questions

**Exercise:**

**Problem:**

Drinking alcoholic beverages causes an increase in urine output. This most likely occurs because alcohol:

- a. inhibits ADH release
- b. stimulates ADH release
- c. inhibits TSH release
- d. stimulates TSH release

---

**Solution:**

A

**Exercise:**

**Problem:**

FSH and LH release from the anterior pituitary is stimulated by \_\_\_\_\_.

- a. TSH
- b. GnRH
- c. T<sub>3</sub>
- d. PTH

---

**Solution:**

B

**Exercise:**

**Problem:** What hormone is produced by beta cells of the pancreas?

- a. T<sub>3</sub>
- b. glucagon
- c. insulin
- d. T<sub>4</sub>

---

**Solution:**

C

**Exercise:**

**Problem:** When blood calcium levels are low, PTH stimulates:

- a. excretion of calcium from the kidneys
- b. excretion of calcium from the intestines
- c. osteoblasts
- d. osteoclasts

---

**Solution:**

D



## Free Response

### Exercise:

#### Problem:

Name and describe a function of one hormone produced by the anterior pituitary and one hormone produced by the posterior pituitary.

---

#### Solution:

In addition to producing FSH and LH, the anterior pituitary also produces the hormone prolactin (PRL) in females. Prolactin stimulates the production of milk by the mammary glands following childbirth. Prolactin levels are regulated by the hypothalamic hormones prolactin-releasing hormone (PRH) and prolactin-inhibiting hormone (PIH) which is now known to be dopamine. PRH stimulates the release of prolactin and PIH inhibits it. The posterior pituitary releases the hormone oxytocin, which stimulates contractions during childbirth. The uterine smooth muscles are not very sensitive to oxytocin until late in pregnancy when the number of oxytocin receptors in the uterus peaks. Stretching of tissues in the uterus and vagina stimulates oxytocin release in childbirth. Contractions increase in intensity as blood levels of oxytocin rise until the birth is complete.

### Exercise:

**Problem:** Describe one direct action of growth hormone (GH).

---

#### Solution:

Hormonal regulation is required for the growth and replication of most cells in the body. Growth hormone (GH), produced by the anterior pituitary, accelerates the rate of protein synthesis, particularly in skeletal muscles and bones. Growth hormone has direct and indirect mechanisms of action. The direct actions of GH include: 1) stimulation of fat breakdown (lipolysis) and release into the blood by adipocytes. This results in a switch by most tissues from utilizing glucose as an

energy source to utilizing fatty acids. This process is called a glucose-sparing effect. 2) In the liver, GH stimulates glycogen breakdown, which is then released into the blood as glucose. Blood glucose levels increase as most tissues are utilizing fatty acids instead of glucose for their energy needs. The GH mediated increase in blood glucose levels is called a diabetogenic effect because it is similar to the high blood glucose levels seen in diabetes mellitus.

## **Glossary**

acromegaly

condition caused by overproduction of GH in adults

Addison's disease

disorder caused by the hyposecretion of corticosteroids

adrenocorticotrophic hormone (ACTH)

hormone released by the anterior pituitary, which stimulates the adrenal cortex to release corticosteroids during the long-term stress response

aldosterone

steroid hormone produced by the adrenal cortex that stimulates the reabsorption of  $\text{Na}^+$  from extracellular fluids and secretion of  $\text{K}^+$ .

androgen

male sex hormone such as testosterone

antidiuretic hormone (ADH)

hormone produced by the hypothalamus and released by the posterior pituitary that increases water reabsorption by the kidneys

calcitonin

hormone produced by the parafollicular cells of the thyroid gland that functions to lower blood  $\text{Ca}^{2+}$  levels and promote bone growth

corticosteroid

hormone released by the adrenal cortex in response to long-term stress

cortisol

glucocorticoid produced in response to stress

Cushing's disease

disorder caused by the hypersecretion of glucocorticoids

diabetes insipidus

disorder caused by underproduction of ADH

diabetes mellitus

disorder caused by low levels of insulin activity

diabetogenic effect

effect of GH that causes blood glucose levels to rise similar to diabetes mellitus

epinephrine

hormone released by the adrenal medulla in response to a short term stress

estrogens

- a group of steroid hormones, including estradiol and several others, that are produced by the ovaries and elicit secondary sex characteristics in females as well as control the maturation of the ova

follicle-stimulating hormone (FSH)

hormone produced by the anterior pituitary that stimulates gamete production

gigantism

condition caused by overproduction of GH in children

glucagon

hormone produced by the alpha cells of the pancreas in response to low blood sugar; functions to raise blood sugar levels

glucocorticoid

corticosteroid that affects glucose metabolism

gluconeogenesis

synthesis of glucose from amino acids

glucose-sparing effect

effect of GH that causes tissues to use fatty acids instead of glucose as an energy source

glycogenolysis

breakdown of glycogen into glucose

goiter

enlargement of the thyroid gland caused by insufficient dietary iodine levels

gonadotropin

hormone that regulates the gonads, including FSH and LH

growth hormone (GH)

hormone produced by the anterior pituitary that promotes protein synthesis and body growth

growth hormone-inhibiting hormone (GHIH)

hormone produced by the hypothalamus that inhibits growth hormone production, also called somatostatin

growth hormone-releasing hormone (GHRH)

hormone released by the hypothalamus that triggers the release of GH

hyperglycemia

high blood sugar level

hyperthyroidism

overactivity of the thyroid gland

hypoglycemia

low blood sugar level

hypothyroidism

underactivity of the thyroid gland

insulin

hormone produced by the beta cells of the pancreas in response to high blood glucose levels; functions to lower blood glucose levels

insulin-like growth factor (IGF)

growth-promoting protein produced by the liver

mineralocorticoid

corticosteroid that affects ion and water balance

norepinephrine

hormone released by the adrenal medulla in response to a short-term stress hormone production by the gonads

osmoreceptor

receptor in the hypothalamus that monitors the concentration of electrolytes in the blood

oxytocin

hormone released by the posterior pituitary to stimulate uterine contractions during childbirth and milk let-down in the mammary glands

parathyroid hormone (PTH)

hormone produced by the parathyroid glands in response to low blood  $\text{Ca}^{2+}$  levels; functions to raise blood  $\text{Ca}^{2+}$  levels

pituitary dwarfism

condition caused by underproduction of GH in children

prolactin (PRL)

hormone produced by the anterior pituitary that stimulates milk production

prolactin-inhibiting hormone

hormone produced by the hypothalamus that inhibits the release of prolactin

prolactin-releasing hormone

hormone produced by the hypothalamus that stimulates the release of prolactin

renin

enzyme produced by the juxtaglomerular apparatus of the kidneys that reacts with angiotensinogen to cause the release of aldosterone

thyroglobulin

glycoprotein found in the thyroid that is converted into thyroid hormone

thyroid-stimulating hormone (TSH)

hormone produced by the anterior pituitary that controls the release of  $T_3$  and  $T_4$  from the thyroid gland

thyroxine (tetraiodothyronine,  $T_4$ )

thyroid hormone containing 4 iodines that controls the basal metabolic rate

triiodothyronine ( $T_3$ )

thyroid hormone containing 3 iodines that controls the basal metabolic rate

## Regulation of Hormone Production

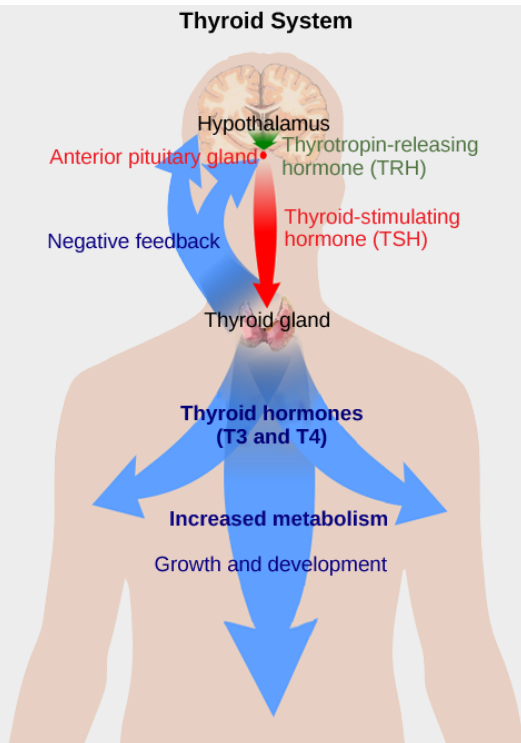
By the end of this section, you will be able to:

- Explain how hormone production is regulated
- Discuss the different stimuli that control hormone levels in the body

Hormone production and release are primarily controlled by negative feedback. In negative feedback systems, a stimulus elicits the release of a substance; once the substance reaches a certain level, it sends a signal that stops further release of the substance. In this way, the concentration of hormones in blood is maintained within a narrow range. For example, the anterior pituitary signals the thyroid to release thyroid hormones. Increasing levels of these hormones in the blood then give feedback to the hypothalamus and anterior pituitary to inhibit further signaling to the thyroid gland, as illustrated in [\[link\]](#). There are three mechanisms by which endocrine glands are stimulated to synthesize and release hormones: humoral stimuli, hormonal stimuli, and neural stimuli.

### **Note:**

Art Connection



The anterior pituitary stimulates the thyroid gland to release thyroid hormones  $T_3$  and  $T_4$ . Increasing levels of these hormones in the blood results in feedback to the hypothalamus and anterior pituitary to inhibit further signaling to the thyroid gland. (credit: modification of work by Mikael Häggström)

Hyperthyroidism is a condition in which the thyroid gland is overactive. Hypothyroidism is a condition in which the thyroid gland is underactive. Which of the conditions are the following two patients most likely to have? Patient A has symptoms including weight gain, cold sensitivity, low heart rate and fatigue.



Patient B has symptoms including weight loss, profuse sweating, increased heart rate and difficulty sleeping.

## Humoral Stimuli

The term “humoral” is derived from the term “humor,” which refers to bodily fluids such as blood. A **humoral stimulus** refers to the control of hormone release in response to changes in extracellular fluids such as blood or the ion concentration in the blood. For example, a rise in blood glucose levels triggers the pancreatic release of insulin. Insulin causes blood glucose levels to drop, which signals the pancreas to stop producing insulin in a negative feedback loop.

## Hormonal Stimuli

**Hormonal stimuli** refers to the release of a hormone in response to another hormone. A number of endocrine glands release hormones when stimulated by hormones released by other endocrine glands. For example, the hypothalamus produces hormones that stimulate the anterior portion of the pituitary gland. The anterior pituitary in turn releases hormones that regulate hormone production by other endocrine glands. The anterior pituitary releases the thyroid-stimulating hormone, which then stimulates the thyroid gland to produce the hormones  $T_3$  and  $T_4$ . As blood concentrations of  $T_3$  and  $T_4$  rise, they inhibit both the pituitary and the hypothalamus in a negative feedback loop.

## Neural Stimuli

In some cases, the nervous system directly stimulates endocrine glands to release hormones, which is referred to as **neural stimuli**. Recall that in a short-term stress response, the hormones epinephrine and norepinephrine are important for providing the bursts of energy required for the body to respond. Here, neuronal signaling from the sympathetic nervous system directly stimulates the adrenal medulla to release the hormones epinephrine and norepinephrine in response to stress.

## Section Summary

Hormone levels are primarily controlled through negative feedback, in which rising levels of a hormone inhibit its further release. The three mechanisms of hormonal release are humoral stimuli, hormonal stimuli, and neural stimuli. Humoral stimuli refers to the control of hormonal release in response to changes in extracellular fluid levels or ion levels. Hormonal stimuli refers to the release of hormones in response to hormones released by other endocrine glands. Neural stimuli refers to the release of hormones in response to neural stimulation.

## Art Connections

### Exercise:

#### Problem:

[\[link\]](#) Hyperthyroidism is a condition in which the thyroid gland is overactive. Hypothyroidism is a condition in which the thyroid gland is underactive. Which of the conditions are the following two patients most likely to have?

Patient A has symptoms including weight gain, cold sensitivity, low heart rate and fatigue.

Patient B has symptoms including weight loss, profuse sweating, increased heart rate and difficulty sleeping.

---

#### Solution:

[\[link\]](#) Patient A has symptoms associated with decreased metabolism, and may be suffering from hypothyroidism. Patient B has symptoms associated with increased metabolism, and may be suffering from hyperthyroidism.

## Review Questions

**Exercise:****Problem:**

A rise in blood glucose levels triggers release of insulin from the pancreas. This mechanism of hormone production is stimulated by:

- a. humoral stimuli
- b. hormonal stimuli
- c. neural stimuli
- d. negative stimuli

---

**Solution:**

A

**Exercise:****Problem:**

Which mechanism of hormonal stimulation would be affected if signaling and hormone release from the hypothalamus was blocked?

- a. humoral and hormonal stimuli
- b. hormonal and neural stimuli
- c. neural and humoral stimuli
- d. hormonal and negative stimuli

---

**Solution:**

B

**Free Response****Exercise:**

**Problem:**

How is hormone production and release primarily controlled?

---

**Solution:**

Hormone production and release are primarily controlled by negative feedback. In negative feedback systems, a stimulus causes the release of a substance whose effects then inhibit further release. In this way, the concentration of hormones in blood is maintained within a narrow range. For example, the anterior pituitary signals the thyroid to release thyroid hormones. Increasing levels of these hormones in the blood then feed back to the hypothalamus and anterior pituitary to inhibit further signaling to the thyroid gland.

**Exercise:**

**Problem:** Compare and contrast hormonal and humoral stimuli.

---

**Solution:**

The term humoral is derived from the term humor, which refers to bodily fluids such as blood. Humoral stimuli refer to the control of hormone release in response to changes in extracellular fluids such as blood or the ion concentration in the blood. For example, a rise in blood glucose levels triggers the pancreatic release of insulin. Insulin causes blood glucose levels to drop, which signals the pancreas to stop producing insulin in a negative feedback loop.

Hormonal stimuli refer to the release of a hormone in response to another hormone. A number of endocrine glands release hormones when stimulated by hormones released by other endocrine organs. For example, the hypothalamus produces hormones that stimulate the anterior pituitary. The anterior pituitary in turn releases hormones that regulate hormone production by other endocrine glands. For example, the anterior pituitary releases thyroid-stimulating hormone, which stimulates the thyroid gland to produce the hormones  $T_3$  and  $T_4$ . As

blood concentrations of  $T_3$  and  $T_4$  rise they inhibit both the pituitary and the hypothalamus in a negative feedback loop.

## **Glossary**

hormonal stimuli

release of a hormone in response to another hormone

humoral stimuli

control of hormone release in response to changes in extracellular fluids such as blood or the ion concentration in the blood

neural stimuli

stimulation of endocrine glands by the nervous system

## Introduction

class="introduction"

Lungs, which appear as nearly transparent tissue surrounding the heart in this X-ray of a dog (left), are the central organs of the respiratory system. The left lung is smaller than the right lung to accommodate space for the heart. A dog's nose (right) has a slit on the side of each nostril. When tracking a scent, the slits open, blocking the front of the nostrils. This allows the dog to exhale

though the  
now-open  
area on the  
side of the  
nostrils  
without  
losing the  
scent that is  
being  
followed.  
(credit a:  
modification  
of work by  
Geoff  
Stearns;  
credit b:  
modification  
of work by  
Cory Zanker)



Breathing is an involuntary event. How often a breath is taken and how much air is inhaled or exhaled are tightly regulated by the respiratory center in the brain. Humans, when they aren't exerting themselves, breathe

approximately 15 times per minute on average. Canines, like the dog in [\[link\]](#), have a respiratory rate of about 15–30 breaths per minute. With every inhalation, air fills the lungs, and with every exhalation, air rushes back out. That air is doing more than just inflating and deflating the lungs in the chest cavity. The air contains oxygen that crosses the lung tissue, enters the bloodstream, and travels to organs and tissues. Oxygen ( $O_2$ ) enters the cells where it is used for metabolic reactions that produce ATP, a high-energy compound. At the same time, these reactions release carbon dioxide ( $CO_2$ ) as a by-product.  $CO_2$  is toxic and must be eliminated. Carbon dioxide exits the cells, enters the bloodstream, travels back to the lungs, and is expired out of the body during exhalation.



## Systems of Gas Exchange

By the end of this section, you will be able to:

- Describe the passage of air from the outside environment to the lungs
- Explain how the lungs are protected from particulate matter

The primary function of the respiratory system is to deliver oxygen to the cells of the body's tissues and remove carbon dioxide, a cell waste product. The main structures of the human respiratory system are the nasal cavity, the trachea, and lungs.

All aerobic organisms require oxygen to carry out their metabolic functions. Along the evolutionary tree, different organisms have devised different means of obtaining oxygen from the surrounding atmosphere. The environment in which the animal lives greatly determines how an animal respire. The complexity of the respiratory system is correlated with the size of the organism. As animal size increases, diffusion distances increase and the ratio of surface area to volume drops. In unicellular organisms, diffusion across the cell membrane is sufficient for supplying oxygen to the cell ([\[link\]](#)). Diffusion is a slow, passive transport process. In order for diffusion to be a feasible means of providing oxygen to the cell, the rate of oxygen uptake must match the rate of diffusion across the membrane. In other words, if the cell were very large or thick, diffusion would not be able to provide oxygen quickly enough to the inside of the cell. Therefore, dependence on diffusion as a means of obtaining oxygen and removing carbon dioxide remains feasible only for small organisms or those with highly-flattened bodies, such as many flatworms (Platyhelminthes). Larger organisms had to evolve specialized respiratory tissues, such as gills, lungs, and respiratory passages accompanied by complex circulatory systems, to transport oxygen throughout their entire body.



The cell of the unicellular algae *Ventricaria ventricosa* is one of the largest known, reaching one to five centimeters in diameter. Like all single-celled organisms, *V. ventricosa* exchanges gases across the cell membrane.

## Direct Diffusion

For small multicellular organisms, diffusion across the outer membrane is sufficient to meet their oxygen needs. Gas exchange by direct diffusion across surface membranes is efficient for organisms less than 1 mm in diameter. In simple organisms, such as cnidarians and flatworms, every cell in the body is close to the external environment. Their cells are kept moist and gases diffuse quickly via direct diffusion. Flatworms are small, literally flat worms, which ‘breathe’ through diffusion across the outer membrane ([link](#)). The flat shape of these organisms increases the surface area for diffusion, ensuring that each cell within the body is close to the outer membrane surface and has access to oxygen. If the flatworm had a

cylindrical body, then the cells in the center would not be able to get oxygen.



This flatworm's process of respiration works by diffusion across the outer membrane. (credit: Stephen Childs)

## Skin and Gills

Earthworms and amphibians use their skin (integument) as a respiratory organ. A dense network of capillaries lies just below the skin and facilitates gas exchange between the external environment and the circulatory system. The respiratory surface must be kept moist in order for the gases to dissolve and diffuse across cell membranes.

Organisms that live in water need to obtain oxygen from the water. Oxygen dissolves in water but at a lower concentration than in the atmosphere. The atmosphere has roughly 21 percent oxygen. In water, the oxygen concentration is much smaller than that. Fish and many other aquatic organisms have evolved gills to take up the dissolved oxygen from water ([link](#)). Gills are thin tissue filaments that are highly branched and folded. When water passes over the gills, the dissolved oxygen in water rapidly

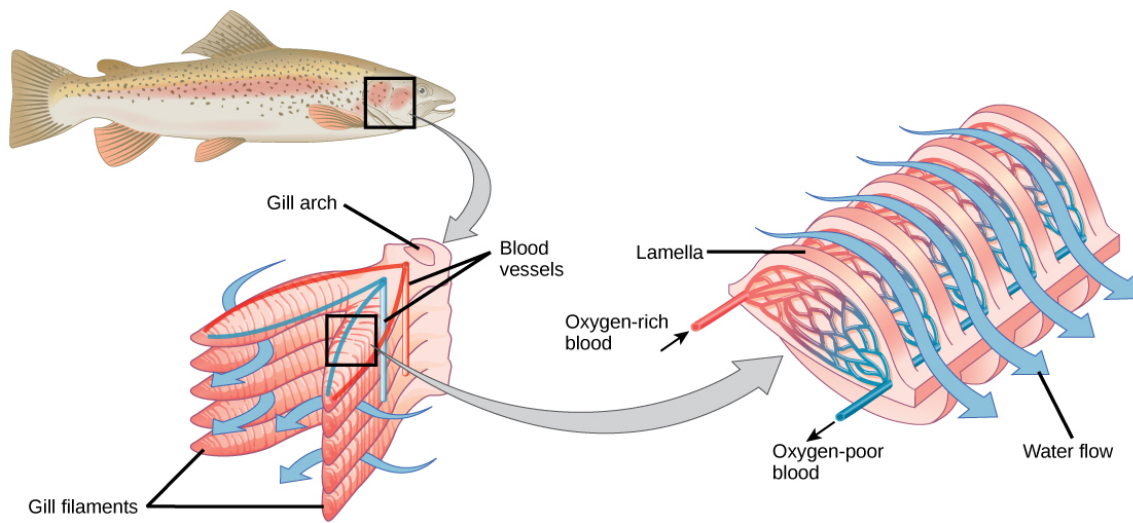
diffuses across the gills into the bloodstream. The circulatory system can then carry the oxygenated blood to the other parts of the body. In animals that contain coelomic fluid instead of blood, oxygen diffuses across the gill surfaces into the coelomic fluid. Gills are found in mollusks, annelids, and crustaceans.



This common carp, like many other aquatic organisms, has gills that allow it to obtain oxygen from water. (credit: "Guitardude012"/Wikimedia Commons)

The folded surfaces of the gills provide a large surface area to ensure that the fish gets sufficient oxygen. Diffusion is a process in which material travels from regions of high concentration to low concentration until equilibrium is reached. In this case, blood with a low concentration of oxygen molecules circulates through the gills. The concentration of oxygen molecules in water is higher than the concentration of oxygen molecules in gills. As a result, oxygen molecules diffuse from water (high concentration) to blood (low concentration), as shown in [\[link\]](#). Similarly, carbon dioxide

molecules in the blood diffuse from the blood (high concentration) to water (low concentration).



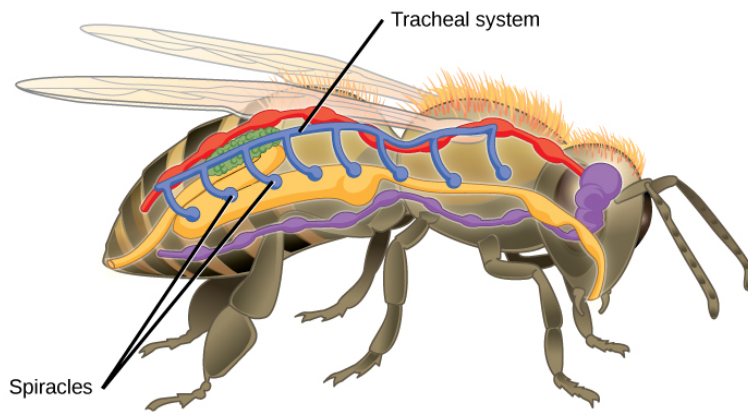
As water flows over the gills, oxygen is transferred to blood via the veins. (credit "fish": modification of work by Duane Raver, NOAA)

## Tracheal Systems

Insect respiration is independent of its circulatory system; therefore, the blood does not play a direct role in oxygen transport. Insects have a highly specialized type of respiratory system called the tracheal system, which consists of a network of small tubes that carries oxygen to the entire body. The tracheal system is the most direct and efficient respiratory system in active animals. The tubes in the tracheal system are made of a polymeric material called chitin.

Insect bodies have openings, called spiracles, along the thorax and abdomen. These openings connect to the tubular network, allowing oxygen to pass into the body ([link](#)) and regulating the diffusion of CO<sub>2</sub> and water

vapor. Air enters and leaves the tracheal system through the spiracles. Some insects can ventilate the tracheal system with body movements.



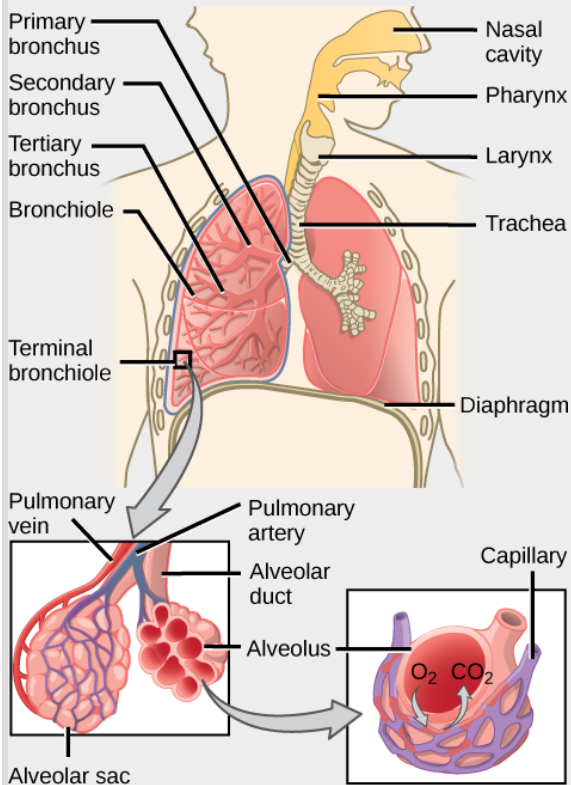
Insects perform respiration via a tracheal system.

## Mammalian Systems

In mammals, pulmonary ventilation occurs via inhalation (breathing). During inhalation, air enters the body through the **nasal cavity** located just inside the nose ([\[link\]](#)). As air passes through the nasal cavity, the air is warmed to body temperature and humidified. The respiratory tract is coated with mucus to seal the tissues from direct contact with air. Mucus is high in water. As air crosses these surfaces of the mucous membranes, it picks up water. These processes help equilibrate the air to the body conditions, reducing any damage that cold, dry air can cause. Particulate matter that is floating in the air is removed in the nasal passages via mucus and cilia. The processes of warming, humidifying, and removing particles are important protective mechanisms that prevent damage to the trachea and lungs. Thus, inhalation serves several purposes in addition to bringing oxygen into the respiratory system.

**Note:**

**Art Connection**



Air enters the respiratory system through the nasal cavity and pharynx, and then passes through the trachea and into the bronchi, which bring air into the lungs.

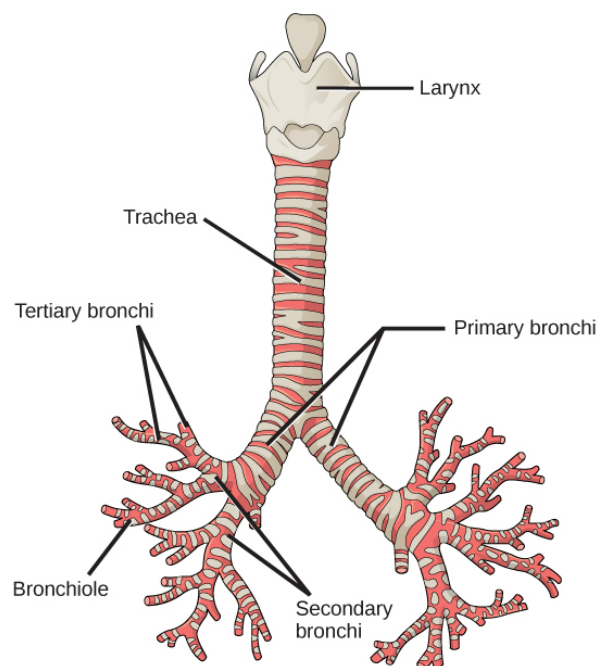
(credit: modification of work by NCI)

Which of the following statements about the mammalian respiratory system is false?

- a. When we breathe in, air travels from the pharynx to the trachea.
- b. The bronchioles branch into bronchi.
- c. Alveolar ducts connect to alveolar sacs.
- d. Gas exchange between the lung and blood takes place in the alveolus.



From the nasal cavity, air passes through the **pharynx** (throat) and the **larynx** (voice box), as it makes its way to the **trachea** ([link](#)). The main function of the trachea is to funnel the inhaled air to the lungs and the exhaled air back out of the body. The human trachea is a cylinder about 10 to 12 cm long and 2 cm in diameter that sits in front of the esophagus and extends from the larynx into the chest cavity where it divides into the two primary bronchi at the midthorax. It is made of incomplete rings of hyaline cartilage and smooth muscle ([link](#)). The trachea is lined with mucus-producing goblet cells and ciliated epithelia. The cilia propel foreign particles trapped in the mucus toward the pharynx. The cartilage provides strength and support to the trachea to keep the passage open. The smooth muscle can contract, decreasing the trachea's diameter, which causes expired air to rush upwards from the lungs at a great force. The forced exhalation helps expel mucus when we cough. Smooth muscle can contract or relax, depending on stimuli from the external environment or the body's nervous system.

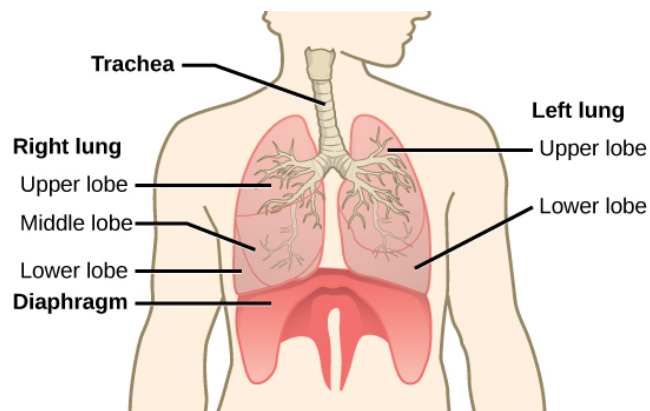




The trachea and bronchi are made of incomplete rings of cartilage. (credit: modification of work by Gray's Anatomy)

## Lungs: Bronchi and Alveoli

The end of the trachea bifurcates (divides) to the right and left lungs. The lungs are not identical. The right lung is larger and contains three lobes, whereas the smaller left lung contains two lobes ([\[link\]](#)). The muscular **diaphragm**, which facilitates breathing, is inferior to (below) the lungs and marks the end of the thoracic cavity.

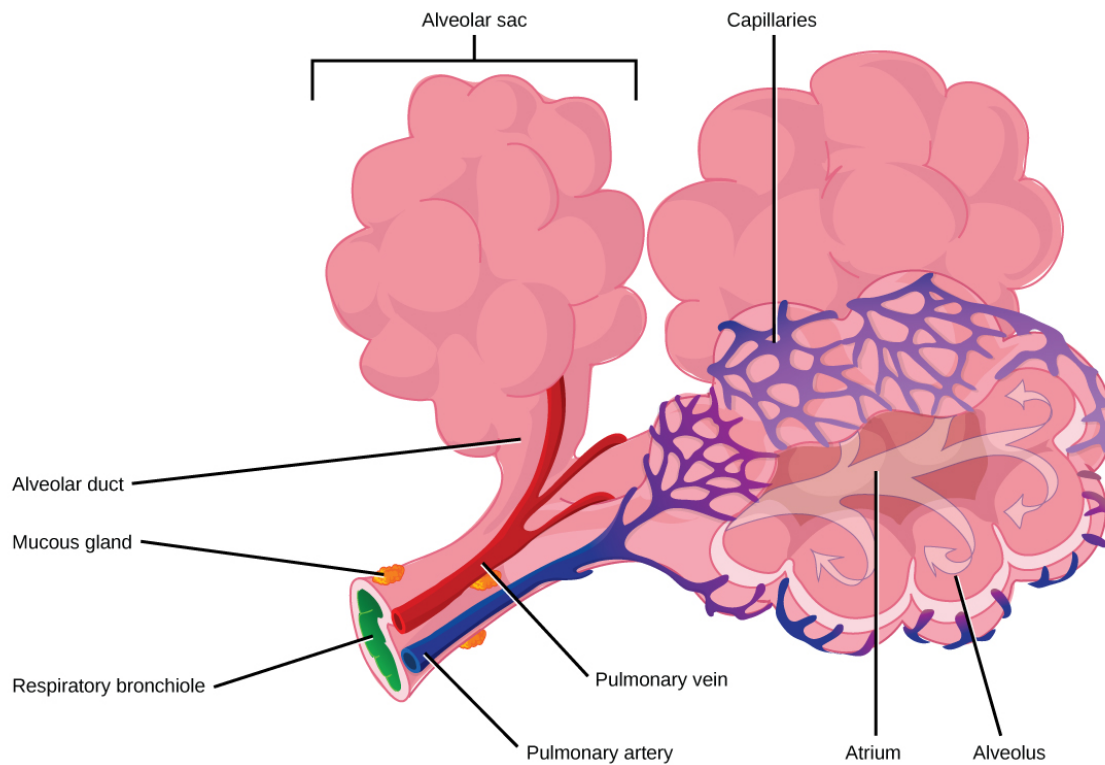


The trachea bifurcates into the right and left bronchi in the lungs. The right lung is made of three lobes and is larger. To accommodate the heart, the left lung is smaller and has only two lobes.

In the lungs, air is diverted into smaller and smaller passages, or **bronchi**. Air enters the lungs through the two **primary (main) bronchi** (singular: bronchus). Each bronchus divides into secondary bronchi, then into tertiary bronchi, which in turn divide, creating smaller and smaller diameter **bronchioles** as they split and spread through the lung. Like the trachea, the bronchi are made of cartilage and smooth muscle. At the bronchioles, the cartilage is replaced with elastic fibers. Bronchi are innervated by nerves of both the parasympathetic and sympathetic nervous systems that control muscle contraction (parasympathetic) or relaxation (sympathetic) in the bronchi and bronchioles, depending on the nervous system's cues. In humans, bronchioles with a diameter smaller than 0.5 mm are the **respiratory bronchioles**. They lack cartilage and therefore rely on inhaled air to support their shape. As the passageways decrease in diameter, the relative amount of smooth muscle increases.

The **terminal bronchioles** subdivide into microscopic branches called respiratory bronchioles. The respiratory bronchioles subdivide into several alveolar ducts. Numerous alveoli and alveolar sacs surround the alveolar ducts. The alveolar sacs resemble bunches of grapes tethered to the end of the bronchioles ([\[link\]](#)). In the acinar region, the **alveolar ducts** are attached to the end of each bronchiole. At the end of each duct are approximately 100 **alveolar sacs**, each containing 20 to 30 **alveoli** that are 200 to 300 microns in diameter. Gas exchange occurs only in alveoli. Alveoli are made of thin-walled parenchymal cells, typically one-cell thick, that look like tiny bubbles within the sacs. Alveoli are in direct contact with capillaries (one-cell thick) of the circulatory system. Such intimate contact ensures that oxygen will diffuse from alveoli into the blood and be distributed to the cells of the body. In addition, the carbon dioxide that was produced by cells as a waste product will diffuse from the blood into alveoli to be exhaled. The anatomical arrangement of capillaries and alveoli emphasizes the structural and functional relationship of the respiratory and circulatory systems. Because there are so many alveoli (~300 million per lung) within each alveolar sac and so many sacs at the end of each alveolar duct, the lungs have a sponge-like consistency. This organization produces a very large surface area that is available for gas exchange. The surface area of alveoli in the lungs is approximately 75 m<sup>2</sup>. This large surface area,

combined with the thin-walled nature of the alveolar parenchymal cells, allows gases to easily diffuse across the cells.



Terminal bronchioles are connected by respiratory bronchioles to alveolar ducts and alveolar sacs. Each alveolar sac contains 20 to 30 spherical alveoli and has the appearance of a bunch of grapes. Air flows into the atrium of the alveolar sac, then circulates into alveoli where gas exchange occurs with the capillaries. Mucous glands secrete mucous into the airways, keeping them moist and flexible. (credit: modification of work by Mariana Ruiz Villareal)

**Note:**

Link to Learning



Watch the following video to review the respiratory system.

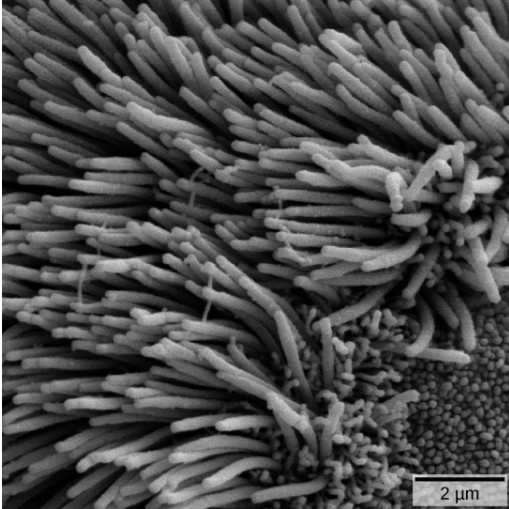
[https://www.openstaxcollege.org/l/lungs\\_pulmonary](https://www.openstaxcollege.org/l/lungs_pulmonary)

## Protective Mechanisms

The air that organisms breathe contains **particulate matter** such as dust, dirt, viral particles, and bacteria that can damage the lungs or trigger allergic immune responses. The respiratory system contains several protective mechanisms to avoid problems or tissue damage. In the nasal cavity, hairs and mucus trap small particles, viruses, bacteria, dust, and dirt to prevent their entry.

If particulates do make it beyond the nose, or enter through the mouth, the bronchi and bronchioles of the lungs also contain several protective devices. The lungs produce **mucus**—a sticky substance made of **mucin**, a complex glycoprotein, as well as salts and water—that traps particulates. The bronchi and bronchioles contain cilia, small hair-like projections that line the walls of the bronchi and bronchioles ([\[link\]](#)). These cilia beat in unison and move mucus and particles out of the bronchi and bronchioles back up to the throat where it is swallowed and eliminated via the esophagus.

In humans, for example, tar and other substances in cigarette smoke destroy or paralyze the cilia, making the removal of particles more difficult. In addition, smoking causes the lungs to produce more mucus, which the damaged cilia are not able to move. This causes a persistent cough, as the lungs try to rid themselves of particulate matter, and makes smokers more susceptible to respiratory ailments.



The bronchi and bronchioles contain cilia that help move mucus and other particles out of the lungs. (credit: Louisa Howard, modification of work by Dartmouth Electron Microscope Facility)

## Section Summary

Animal respiratory systems are designed to facilitate gas exchange. In mammals, air is warmed and humidified in the nasal cavity. Air then travels down the pharynx, through the trachea, and into the lungs. In the lungs, air passes through the branching bronchi, reaching the respiratory bronchioles, which house the first site of gas exchange. The respiratory bronchioles open into the alveolar ducts, alveolar sacs, and alveoli. Because there are so many alveoli and alveolar sacs in the lung, the surface area for gas exchange is very large. Several protective mechanisms are in place to prevent damage or infection. These include the hair and mucus in the nasal cavity that trap dust, dirt, and other particulate matter before they can enter

the system. In the lungs, particles are trapped in a mucus layer and transported via cilia up to the esophageal opening at the top of the trachea to be swallowed.

**Exercise:**

**Problem:**

[\[link\]](#) Which of the following statements about the mammalian respiratory system is false?

- a. When we breathe in, air travels from the pharynx to the trachea.
- b. The bronchioles branch into bronchi.
- c. Alveolar ducts connect to alveolar sacs.
- d. Gas exchange between the lung and blood takes place in the alveolus.

---

**Solution:**

[\[link\]](#) B

## Review Questions

**Exercise:**

**Problem:** The respiratory system \_\_\_\_\_.

- a. provides body tissues with oxygen
- b. provides body tissues with oxygen and carbon dioxide
- c. establishes how many breaths are taken per minute
- d. provides the body with carbon dioxide

---

**Solution:**

A

**Exercise:**

**Problem:**

Air is warmed and humidified in the nasal passages. This helps to \_\_\_\_\_.

- a. ward off infection
- b. decrease sensitivity during breathing
- c. prevent damage to the lungs
- d. all of the above

---

**Solution:**

C

**Exercise:**

**Problem:** Which is the order of airflow during inhalation?

- a. nasal cavity, trachea, larynx, bronchi, bronchioles, alveoli
- b. nasal cavity, larynx, trachea, bronchi, bronchioles, alveoli
- c. nasal cavity, larynx, trachea, bronchioles, bronchi, alveoli
- d. nasal cavity, trachea, larynx, bronchi, bronchioles, alveoli

---

**Solution:**

B

**Free Response****Exercise:****Problem:**

Describe the function of these terms and describe where they are located: main bronchus, trachea, alveoli, and acinus.

---

**Solution:**

The main bronchus is the conduit in the lung that funnels air to the airways where gas exchange occurs. The main bronchus attaches the lungs to the very end of the trachea where it bifurcates. The trachea is the cartilaginous structure that extends from the pharynx to the primary bronchi. It serves to funnel air to the lungs. The alveoli are the sites of gas exchange; they are located at the terminal regions of the lung and are attached to the respiratory bronchioles. The acinus is the structure in the lung where gas exchange occurs.

**Exercise:**

**Problem:**How does the structure of alveoli maximize gas exchange?

---

**Solution:**

The sac-like structure of the alveoli increases their surface area. In addition, the alveoli are made of thin-walled parenchymal cells. These features allow gases to easily diffuse across the cells.

**Glossary**

alveolar duct

duct that extends from the terminal bronchiole to the alveolar sac

alveolar sac

structure consisting of two or more alveoli that share a common opening

alveolus

(plural: alveoli) (also, air sac) terminal region of the lung where gas exchange occurs

bronchus

(plural: bronchi) smaller branch of cartilaginous tissue that stems off of the trachea; air is funneled through the bronchi to the region where gas



exchange occurs in alveoli

bronchiole

airway that extends from the main tertiary bronchi to the alveolar sac

diaphragm

domed-shaped skeletal muscle located under lungs that separates the thoracic cavity from the abdominal cavity

larynx

voice box, a short passageway connecting the pharynx and the trachea

mucin

complex glycoprotein found in mucus

mucus

sticky protein-containing fluid secretion in the lung that traps particulate matter to be expelled from the body

nasal cavity

opening of the respiratory system to the outside environment

particulate matter

small particle such as dust, dirt, viral particles, and bacteria that are in the air

pharynx

throat; a tube that starts in the internal nares and runs partway down the neck, where it opens into the esophagus and the larynx

primary bronchus

(also, main bronchus) region of the airway within the lung that attaches to the trachea and bifurcates to each lung where it branches into secondary bronchi

respiratory bronchiole

terminal portion of the bronchiole tree that is attached to the terminal bronchioles and alveoli ducts, alveolar sacs, and alveoli

terminal bronchiole

region of bronchiole that attaches to the respiratory bronchioles

trachea

cartilaginous tube that transports air from the larynx to the primary bronchi

## Gas Exchange across Respiratory Surfaces

By the end of this section, you will be able to:

- Name and describe lung volumes and capacities
- Understand how gas pressure influences how gases move into and out of the body

The structure of the lung maximizes its surface area to increase gas diffusion. Because of the enormous number of alveoli (approximately 300 million in each human lung), the surface area of the lung is very large ( $75 \text{ m}^2$ ). Having such a large surface area increases the amount of gas that can diffuse into and out of the lungs.

## Basic Principles of Gas Exchange

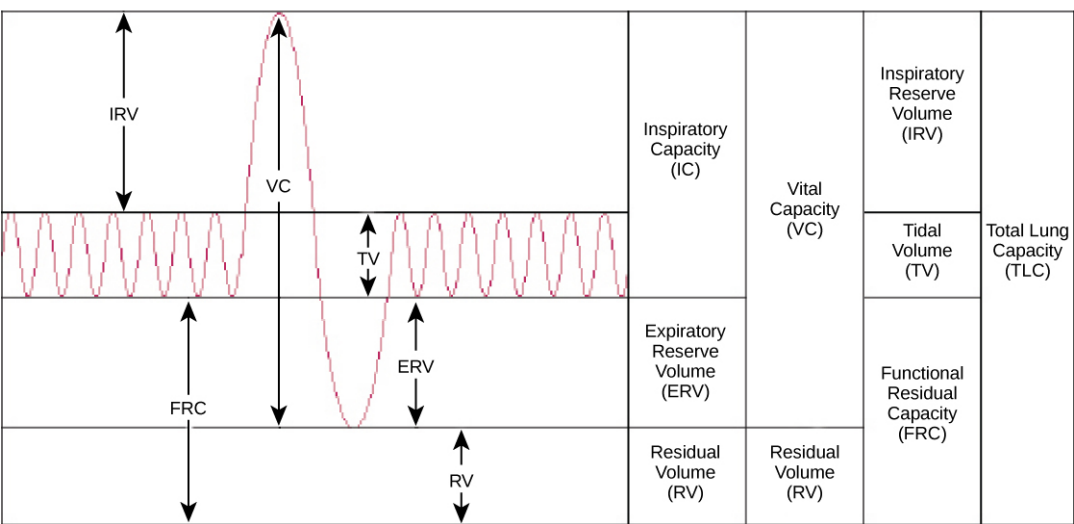
Gas exchange during respiration occurs primarily through diffusion. Diffusion is a process in which transport is driven by a concentration gradient. Gas molecules move from a region of high concentration to a region of low concentration. Blood that is low in oxygen concentration and high in carbon dioxide concentration undergoes gas exchange with air in the lungs. The air in the lungs has a higher concentration of oxygen than that of oxygen-depleted blood and a lower concentration of carbon dioxide. This concentration gradient allows for gas exchange during respiration.

**Partial pressure** is a measure of the concentration of the individual components in a mixture of gases. The total pressure exerted by the mixture is the sum of the partial pressures of the components in the mixture. The rate of diffusion of a gas is proportional to its partial pressure within the total gas mixture. This concept is discussed further in detail below.

## Lung Volumes and Capacities

Different animals have different lung capacities based on their activities. Cheetahs have evolved a much higher lung capacity than humans; it helps provide oxygen to all the muscles in the body and allows them to run very fast. Elephants also have a high lung capacity. In this case, it is not because they run fast but because they have a large body and must be able to take up oxygen in accordance with their body size.

Human lung size is determined by genetics, sex, and height. At maximal capacity, an average lung can hold almost six liters of air, but lungs do not usually operate at maximal capacity. Air in the lungs is measured in terms of **lung volumes** and **lung capacities** ([\[link\]](#) and [\[link\]](#)). Volume measures the amount of air for one function (such as inhalation or exhalation). Capacity is any two or more volumes (for example, how much can be inhaled from the end of a maximal exhalation).



Human lung volumes and capacities are shown. The total lung capacity of the adult male is six liters. Tidal volume is the volume of air inhaled in a single, normal breath. Inspiratory capacity is the amount of air taken in during a deep breath, and residual volume is the amount of air left in the lungs after forceful respiration.

**Lung Volumes and Capacities (Avg Adult Male)**

---

<b>Lung Volumes and Capacities (Avg Adult Male)</b>			
<b>Volume/Capacity</b>	<b>Definition</b>	<b>Volume (liters)</b>	<b>Equations</b>
<b>Volume/Capacity</b>	<b>Definition</b>	<b>Volume (liters)</b>	<b>Equations</b>
Tidal volume (TV)	Amount of air inhaled during a normal breath	0.5	-
Expiratory reserve volume (ERV)	Amount of air that can be exhaled after a normal exhalation	1.2	-
Inspiratory reserve volume (IRV)	Amount of air that can be further inhaled after a normal inhalation	3.1	-
Residual volume (RV)	Air left in the lungs after a forced exhalation	1.2	-

### **Lung Volumes and Capacities (Avg Adult Male)**

<b>Volume/Capacity</b>	<b>Definition</b>	<b>Volume (liters)</b>	<b>Equations</b>
Vital capacity (VC)	Maximum amount of air that can be moved in or out of the lungs in a single respiratory cycle	4.8	$ERV + TV + IRV$
Inspiratory capacity (IC)	Volume of air that can be inhaled in addition to a normal exhalation	3.6	$TV + IRV$
Functional residual capacity (FRC)	Volume of air remaining after a normal exhalation	2.4	$ERV + RV$

### Lung Volumes and Capacities (Avg Adult Male)

Volume/Capacity	Definition	Volume (liters)	Equations
Total lung capacity (TLC)	Total volume of air in the lungs after a maximal inspiration	6.0	$RV + ERV + TV + IRV$
Forced expiratory volume (FEV1)	How much air can be forced out of the lungs over a specific time period, usually one second	~4.1 to 5.5	-

The volume in the lung can be divided into four units: tidal volume, expiratory reserve volume, inspiratory reserve volume, and residual volume. **Tidal volume (TV)** measures the amount of air that is inspired and expired during a normal breath. On average, this volume is around one-half liter, which is a little less than the capacity of a 20-ounce drink bottle. The **expiratory reserve volume (ERV)** is the additional amount of air that can be exhaled after a normal exhalation. It is the reserve amount that can be exhaled beyond what is normal. Conversely, the **inspiratory reserve volume (IRV)** is the additional amount of air that can be inhaled after a normal inhalation. The **residual volume (RV)** is the amount of air that is left after expiratory reserve volume is exhaled. The lungs are never completely empty: There is always

some air left in the lungs after a maximal exhalation. If this residual volume did not exist and the lungs emptied completely, the lung tissues would stick together and the energy necessary to re-inflate the lung could be too great to overcome. Therefore, there is always some air remaining in the lungs. Residual volume is also important for preventing large fluctuations in respiratory gases ( $O_2$  and  $CO_2$ ). The residual volume is the only lung volume that cannot be measured directly because it is impossible to completely empty the lung of air. This volume can only be calculated rather than measured.

Capacities are measurements of two or more volumes. The **vital capacity (VC)** measures the maximum amount of air that can be inhaled or exhaled during a respiratory cycle. It is the sum of the expiratory reserve volume, tidal volume, and inspiratory reserve volume. The **inspiratory capacity (IC)** is the amount of air that can be inhaled after the end of a normal expiration. It is, therefore, the sum of the tidal volume and inspiratory reserve volume. The **functional residual capacity (FRC)** includes the expiratory reserve volume and the residual volume. The FRC measures the amount of additional air that can be exhaled after a normal exhalation. Lastly, the **total lung capacity (TLC)** is a measurement of the total amount of air that the lung can hold. It is the sum of the residual volume, expiratory reserve volume, tidal volume, and inspiratory reserve volume.

Lung volumes are measured by a technique called **spirometry**. An important measurement taken during spirometry is the **forced expiratory volume (FEV)**, which measures how much air can be forced out of the lung over a specific period, usually one second (FEV1). In addition, the forced vital capacity (FVC), which is the total amount of air that can be forcibly exhaled, is measured. The ratio of these values (**FEV1/FVC ratio**) is used to diagnose lung diseases including asthma, emphysema, and fibrosis. If the FEV1/FVC ratio is high, the lungs are not compliant (meaning they are stiff and unable to bend properly), and the patient most likely has lung fibrosis. Patients exhale most of the lung volume very quickly. Conversely, when the FEV1/FVC ratio is low, there is resistance in the lung that is characteristic of asthma. In this instance, it is hard for the patient to get the air out of his or her lungs, and it takes a long time to reach the maximal exhalation volume. In either case, breathing is difficult and complications arise.



**Note:****Career Connection****Respiratory Therapist**

Respiratory therapists or respiratory practitioners evaluate and treat patients with lung and cardiovascular diseases. They work as part of a medical team to develop treatment plans for patients. Respiratory therapists may treat premature babies with underdeveloped lungs, patients with chronic conditions such as asthma, or older patients suffering from lung disease such as emphysema and chronic obstructive pulmonary disease (COPD). They may operate advanced equipment such as compressed gas delivery systems, ventilators, blood gas analyzers, and resuscitators. Specialized programs to become a respiratory therapist generally lead to a bachelor's degree with a respiratory therapist specialty. Because of a growing aging population, career opportunities as a respiratory therapist are expected to remain strong.

## Gas Pressure and Respiration

The respiratory process can be better understood by examining the properties of gases. Gases move freely, but gas particles are constantly hitting the walls of their vessel, thereby producing gas pressure.

Air is a mixture of gases, primarily nitrogen ( $N_2$ ; 78.6 percent), oxygen ( $O_2$ ; 20.9 percent), water vapor ( $H_2O$ ; 0.5 percent), and carbon dioxide ( $CO_2$ ; 0.04 percent). Each gas component of that mixture exerts a pressure. The pressure for an individual gas in the mixture is the partial pressure of that gas.

Approximately 21 percent of atmospheric gas is oxygen. Carbon dioxide, however, is found in relatively small amounts, 0.04 percent. The partial pressure for oxygen is much greater than that of carbon dioxide. The partial pressure of any gas can be calculated by:

**Equation:**

$$P = (P_{\text{atm}}) \times (\text{percent content in mixture}).$$

$P_{\text{atm}}$ , the atmospheric pressure, is the sum of all of the partial pressures of the atmospheric gases added together,

**Equation:**

$$P_{\text{atm}} = P_{\text{N}_2} + P_{\text{O}_2} + P_{\text{H}_2\text{O}} + P_{\text{CO}_2} = 760 \text{ mm Hg}$$

× (percent content in mixture).

The pressure of the atmosphere at sea level is 760 mm Hg. Therefore, the partial pressure of oxygen is:

**Equation:**

$$P_{\text{O}_2} = (760 \text{ mm Hg}) (0.21) = 160 \text{ mm Hg}$$

and for carbon dioxide:

**Equation:**

$$P_{\text{CO}_2} = (760 \text{ mm Hg}) (0.0004) = 0.3 \text{ mm Hg}.$$

At high altitudes,  $P_{\text{atm}}$  decreases but concentration does not change; the partial pressure decrease is due to the reduction in  $P_{\text{atm}}$ .

When the air mixture reaches the lung, it has been humidified. The pressure of the water vapor in the lung does not change the pressure of the air, but it must be included in the partial pressure equation. For this calculation, the water pressure (47 mm Hg) is subtracted from the atmospheric pressure:

**Equation:**

$$760 \text{ mm Hg} - 47 \text{ mm Hg} = 713 \text{ mm Hg}$$

and the partial pressure of oxygen is:

**Equation:**

$$(760 \text{ mm Hg} - 47 \text{ mm Hg}) \times 0.21 = 150 \text{ mm Hg}.$$

These pressures determine the gas exchange, or the flow of gas, in the system. Oxygen and carbon dioxide will flow according to their pressure gradient from high to low. Therefore, understanding the partial pressure of each gas will aid in understanding how gases move in the respiratory system.

## Gas Exchange across the Alveoli

In the body, oxygen is used by cells of the body's tissues and carbon dioxide is produced as a waste product. The ratio of carbon dioxide production to oxygen consumption is the **respiratory quotient (RQ)**. RQ varies between 0.7 and 1.0. If just glucose were used to fuel the body, the RQ would equal one. One mole of carbon dioxide would be produced for every mole of oxygen consumed. Glucose, however, is not the only fuel for the body. Protein and fat are also used as fuels for the body. Because of this, less carbon dioxide is produced than oxygen is consumed and the RQ is, on average, about 0.7 for fat and about 0.8 for protein.

The RQ is used to calculate the partial pressure of oxygen in the alveolar spaces within the lung, the **alveolar**  $P_{O_2}$ . Above, the partial pressure of oxygen in the lungs was calculated to be 150 mm Hg. However, lungs never fully deflate with an exhalation; therefore, the inspired air mixes with this residual air and lowers the partial pressure of oxygen within the alveoli. This means that there is a lower concentration of oxygen in the lungs than is found in the air outside the body. Knowing the RQ, the partial pressure of oxygen in the alveoli can be calculated:

**Equation:**

$$\text{alveolar } P_{O_2} = \text{inspired } P_{O_2} - \left( \frac{\text{alveolar } P_{CO_2}}{RQ} \right)$$

With an RQ of 0.8 and a  $P_{CO_2}$  in the alveoli of 40 mm Hg, the alveolar  $P_{O_2}$  is equal to:

**Equation:**

$$\text{alveolar } P_{O_2} = 150 \text{ mm Hg} - \left( \frac{40 \text{ mm Hg}}{0.8} \right) = \text{mm Hg.}$$

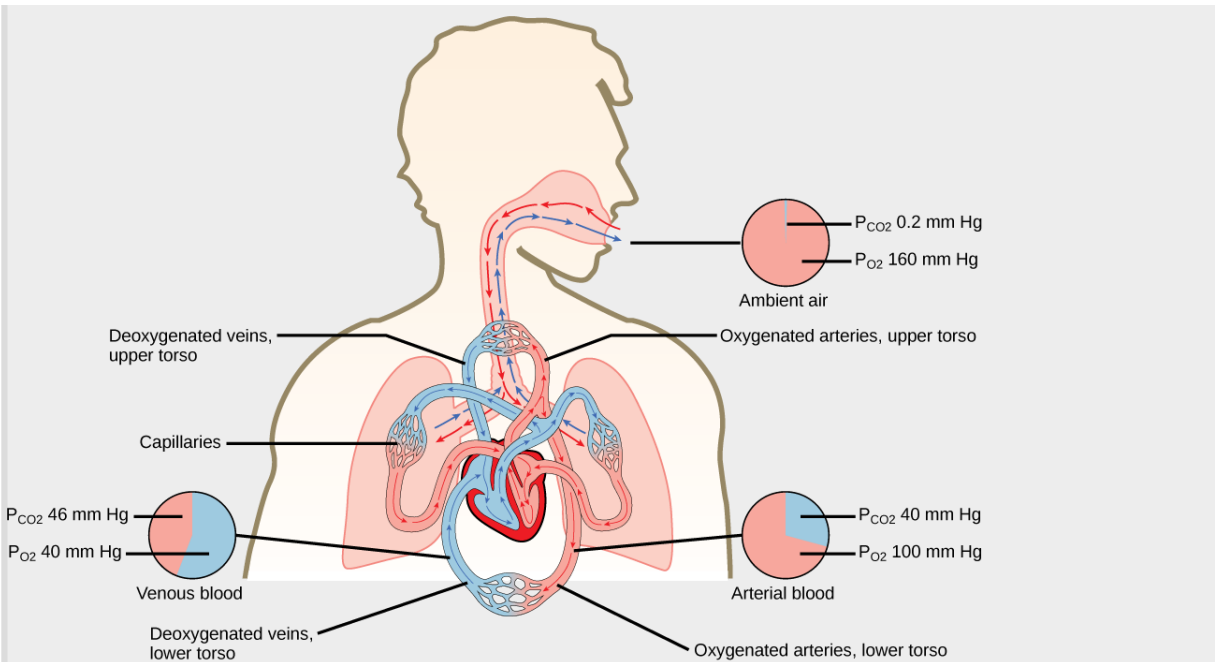
Notice that this pressure is less than the external air. Therefore, the oxygen will flow from the inspired air in the lung ( $P_{O_2} = 150 \text{ mm Hg}$ ) into the bloodstream ( $P_{O_2} = 100 \text{ mm Hg}$ ) ([link](#)).

In the lungs, oxygen diffuses out of the alveoli and into the capillaries surrounding the alveoli. Oxygen (about 98 percent) binds reversibly to the respiratory pigment hemoglobin found in red blood cells (RBCs). RBCs carry oxygen to the tissues where oxygen dissociates from the hemoglobin and diffuses into the cells of the tissues. More specifically, alveolar  $P_{O_2}$  is higher in the alveoli ( $P_{ALVO_2} = 100$  mm Hg) than blood  $P_{O_2}$  (40 mm Hg) in the capillaries. Because this pressure gradient exists, oxygen diffuses down its pressure gradient, moving out of the alveoli and entering the blood of the capillaries where  $O_2$  binds to hemoglobin. At the same time, alveolar  $P_{CO_2}$  is lower  $P_{ALVO_2} = 40$  mm Hg than blood  $P_{CO_2} = (45$  mm Hg).  $CO_2$  diffuses down its pressure gradient, moving out of the capillaries and entering the alveoli.

Oxygen and carbon dioxide move independently of each other; they diffuse down their own pressure gradients. As blood leaves the lungs through the pulmonary veins, the **venous**  $P_{O_2} = 100$  mm Hg, whereas the **venous**  $P_{CO_2} = 40$  mm Hg. As blood enters the systemic capillaries, the blood will lose oxygen and gain carbon dioxide because of the pressure difference of the tissues and blood. In systemic capillaries,  $P_{O_2} = 100$  mm Hg, but in the tissue cells,  $P_{O_2} = 40$  mm Hg. This pressure gradient drives the diffusion of oxygen out of the capillaries and into the tissue cells. At the same time, blood  $P_{CO_2} = 40$  mm Hg and systemic tissue  $P_{CO_2} = 45$  mm Hg. The pressure gradient drives  $CO_2$  out of tissue cells and into the capillaries. The blood returning to the lungs through the pulmonary arteries has a venous  $P_{O_2} = 40$  mm Hg and a  $P_{CO_2} = 45$  mm Hg. The blood enters the lung capillaries where the process of exchanging gases between the capillaries and alveoli begins again ([\[link\]](#)).

**Note:**

Art Connection



The partial pressures of oxygen and carbon dioxide change as blood moves through the body.

Which of the following statements is false?

- In the tissues,  $P_{O_2}$  drops as blood passes from the arteries to the veins, while  $P_{CO_2}$  increases.
- Blood travels from the lungs to the heart to body tissues, then back to the heart, then the lungs.
- Blood travels from the lungs to the heart to body tissues, then back to the lungs, then the heart.
- $P_{O_2}$  is higher in air than in the lungs.

In short, the change in partial pressure from the alveoli to the capillaries drives the oxygen into the tissues and the carbon dioxide into the blood from the tissues. The blood is then transported to the lungs where differences in pressure in the alveoli result in the movement of carbon dioxide out of the blood into the lungs, and oxygen into the blood.

**Note:****Link to Learning**

Watch this video to learn how to carry out spirometry.

<https://www.openstaxcollege.org/l/spirometry>

## Section Summary

The lungs can hold a large volume of air, but they are not usually filled to maximal capacity. Lung volume measurements include tidal volume, expiratory reserve volume, inspiratory reserve volume, and residual volume. The sum of these equals the total lung capacity. Gas movement into or out of the lungs is dependent on the pressure of the gas. Air is a mixture of gases; therefore, the partial pressure of each gas can be calculated to determine how the gas will flow in the lung. The difference between the partial pressure of the gas in the air drives oxygen into the tissues and carbon dioxide out of the body.

## Art Connections

**Exercise:**

**Problem:** [\[link\]](#) Which of the following statements is false?

- In the tissues,  $P_{O_2}$  drops as blood passes from the arteries to the veins, while  $P_{CO_2}$  increases.
- Blood travels from the lungs to the heart to body tissues, then back to the heart, then the lungs.
- Blood travels from the lungs to the heart to body tissues, then back to the lungs, then the heart.

d.  $P_{O_2}$  is higher in air than in the lungs.

---

**Solution:**

[\[link\]](#) C

## Review Questions

**Exercise:**

**Problem:** The inspiratory reserve volume measures the \_\_\_\_\_.

- a. amount of air remaining in the lung after a maximal exhalation
  - b. amount of air that the lung holds
  - c. amount of air that can be further exhaled after a normal breath
  - d. amount of air that can be further inhaled after a normal breath
- 

**Solution:**

D

**Exercise:**

**Problem:**

Of the following, which does not explain why the partial pressure of oxygen is lower in the lung than in the external air?

- a. Air in the lung is humidified; therefore, water vapor pressure alters the pressure.
  - b. Carbon dioxide mixes with oxygen.
  - c. Oxygen is moved into the blood and is headed to the tissues.
  - d. Lungs exert a pressure on the air to reduce the oxygen pressure.
- 

**Solution:**

D

**Exercise:**

**Problem:**

The total lung capacity is calculated using which of the following formulas?

- a. residual volume + tidal volume + inspiratory reserve volume
- b. residual volume + expiratory reserve volume + inspiratory reserve volume
- c. expiratory reserve volume + tidal volume + inspiratory reserve volume
- d. residual volume + expiratory reserve volume + tidal volume + inspiratory reserve volume

---

**Solution:**

D

**Free Response**

**Exercise:**

**Problem:**

What does FEV1/FVC measure? What factors may affect FEV1/FVC?

---

**Solution:**

FEV1/FVC measures the forced expiratory volume in one second in relation to the total forced vital capacity (the total amount of air that is exhaled from the lung from a maximal inhalation). This ratio changes with alterations in lung function that arise from diseases such as fibrosis, asthma, and COPD.

**Exercise:**



**Problem:** What is the reason for having residual volume in the lung?

---

**Solution:**

If all the air in the lung were exhaled, then opening the alveoli for the next inspiration would be very difficult. This is because the tissues would stick together.

**Exercise:**

**Problem:**

How can a decrease in the percent of oxygen in the air affect the movement of oxygen in the body?

---

**Solution:**

Oxygen moves from the lung to the bloodstream to the tissues according to the pressure gradient. This is measured as the partial pressure of oxygen. If the amount of oxygen drops in the inspired air, there would be reduced partial pressure. This would decrease the driving force that moves the oxygen into the blood and into the tissues.  $P_{O_2}$  is also reduced at high elevations:  $P_{O_2}$  at high elevations is lower than at sea level because the total atmospheric pressure is less than atmospheric pressure at sea level.

**Exercise:**

**Problem:**

If a patient has increased resistance in his or her lungs, how can this be detected by a doctor? What does this mean?

---

**Solution:**

A doctor can detect a restrictive disease using spirometry. By detecting the rate at which air can be expelled from the lung, a diagnosis of fibrosis or another restrictive disease can be made.

## Glossary

alveolar  $P_{O_2}$

partial pressure of oxygen in the alveoli (usually around 100 mmHg)

expiratory reserve volume (ERV)

amount of additional air that can be exhaled after a normal exhalation

FEV1/FVC ratio

ratio of how much air can be forced out of the lung in one second to the total amount that is forced out of the lung; a measurement of lung function that can be used to detect disease states

forced expiratory volume (FEV)

(also, forced vital capacity) measure of how much air can be forced out of the lung from maximal inspiration over a specific amount of time

functional residual capacity (FRC)

expiratory reserve volume plus residual volume

inspiratory capacity (IC)

tidal volume plus inspiratory reserve volume

inspiratory reserve volume (IRV)

amount of additional air that can be inspired after a normal inhalation

lung capacity

measurement of two or more lung volumes (how much air can be inhaled from the end of an expiration to maximal capacity)

lung volume

measurement of air for one lung function (normal inhalation or exhalation)

partial pressure

amount of pressure exerted by one gas within a mixture of gases

residual volume (RV)

amount of air remaining in the lung after a maximal expiration

respiratory quotient (RQ)

ratio of carbon dioxide production to each oxygen molecule consumed

spirometry

method to measure lung volumes and to diagnose lung diseases

tidal volume (TV)

amount of air that is inspired and expired during normal breathing

total lung capacity (TLC)

sum of the residual volume, expiratory reserve volume, tidal volume, and inspiratory reserve volume

venous  $P_{CO_2}$

partial pressure of carbon dioxide in the veins (40 mm Hg in the pulmonary veins)

venous  $P_{O_2}$

partial pressure of oxygen in the veins (100 mm Hg in the pulmonary veins)

**vital capacity (VC)**

sum of the expiratory reserve volume, tidal volume, and inspiratory reserve volume

## Breathing

By the end of this section, you will be able to:

- Describe how the structures of the lungs and thoracic cavity control the mechanics of breathing
- Explain the importance of compliance and resistance in the lungs
- Discuss problems that may arise due to a V/Q mismatch

Mammalian lungs are located in the thoracic cavity where they are surrounded and protected by the rib cage, intercostal muscles, and bound by the chest wall. The bottom of the lungs is contained by the diaphragm, a skeletal muscle that facilitates breathing. Breathing requires the coordination of the lungs, the chest wall, and most importantly, the diaphragm.

## Types of Breathing

Amphibians have evolved multiple ways of breathing. Young amphibians, like tadpoles, use gills to breathe, and they don't leave the water. Some amphibians retain gills for life. As the tadpole grows, the gills disappear and lungs grow. These lungs are primitive and not as evolved as mammalian lungs. Adult amphibians are lacking or have a reduced diaphragm, so breathing via lungs is forced. The other means of breathing for amphibians is diffusion across the skin. To aid this diffusion, amphibian skin must remain moist.

Birds face a unique challenge with respect to breathing: They fly. Flying consumes a great amount of energy; therefore, birds require a lot of oxygen to aid their metabolic processes. Birds have evolved a respiratory system that supplies them with the oxygen needed to enable flying. Similar to mammals, birds have lungs, which are organs specialized for gas exchange. Oxygenated air, taken in during inhalation, diffuses across the surface of the lungs into the bloodstream, and carbon dioxide diffuses from the blood into the lungs and expelled during exhalation. The details of breathing between birds and mammals differ substantially.

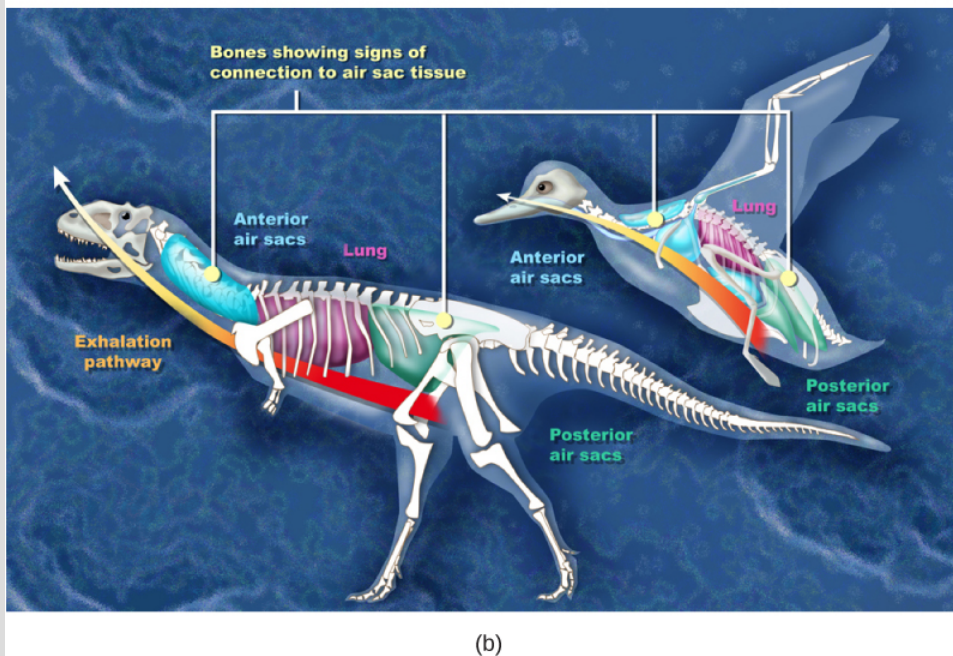
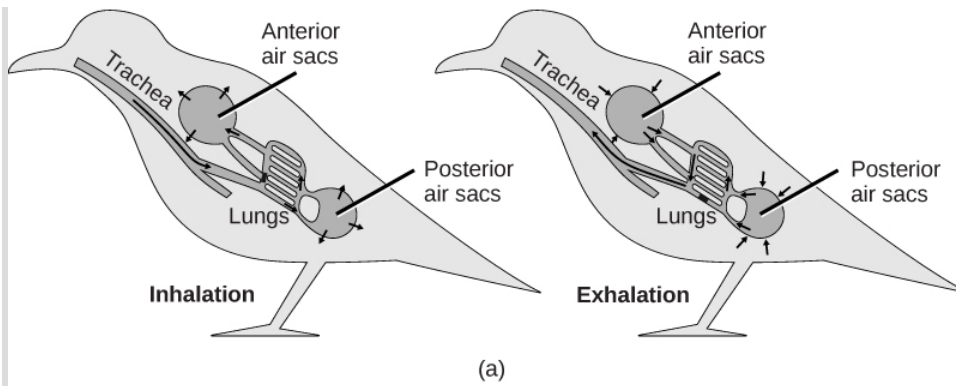
In addition to lungs, birds have air sacs inside their body. Air flows in one direction from the posterior air sacs to the lungs and out of the anterior air sacs. The flow of air is in the opposite direction from blood flow, and gas exchange takes place much more efficiently. This type of breathing enables birds to obtain the requisite oxygen, even at higher altitudes where the oxygen concentration is low. This directionality of airflow requires two cycles of air intake and exhalation to completely get the air out of the lungs.

**Note:****Evolution Connection****Avian Respiration**

Birds have evolved a respiratory system that enables them to fly. Flying is a high-energy process and requires a lot of oxygen. Furthermore, many birds fly in high altitudes where the concentration of oxygen is low. How did birds evolve a respiratory system that is so unique?

Decades of research by paleontologists have shown that birds evolved from theropods, meat-eating dinosaurs ([\[link\]](#)). In fact, fossil evidence shows that meat-eating dinosaurs that lived more than 100 million years ago had a similar flow-through respiratory system with lungs and air sacs.

*Archaeopteryx* and *Xiaotingia*, for example, were flying dinosaurs and are believed to be early precursors of birds.



(a) Birds have a flow-through respiratory system in which air flows unidirectionally from the posterior sacs into the lungs, then into the anterior air sacs. The air sacs connect to openings in hollow bones. (b) Dinosaurs, from which birds descended, have similar hollow bones and are believed to have had a similar respiratory system. (credit b: modification of work by Zina Deretsky, National Science Foundation)

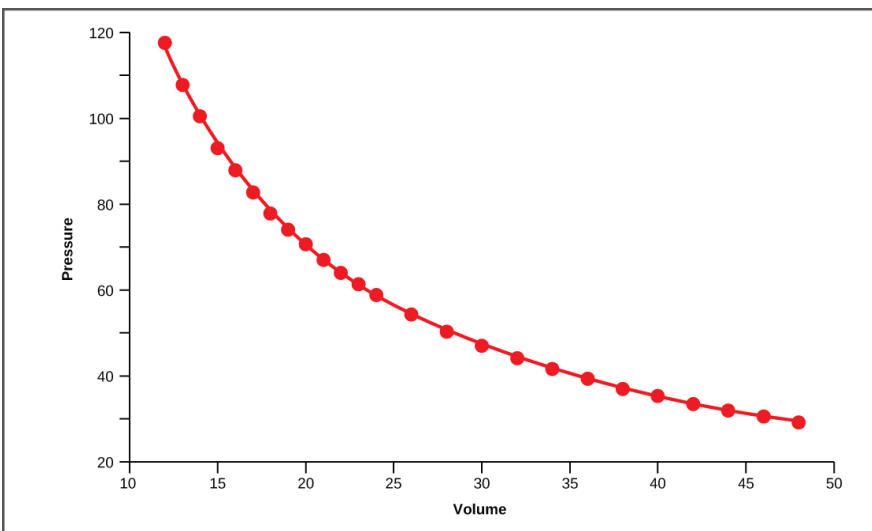
Most of us consider that dinosaurs are extinct. However, modern birds are descendants of avian dinosaurs. The respiratory system of modern birds has been evolving for hundreds of millions of years.

---

All mammals have lungs that are the main organs for breathing. Lung capacity has evolved to support the animal's activities. During inhalation, the lungs expand with air, and oxygen diffuses across the lung's surface and enters the bloodstream. During exhalation, the lungs expel air and lung volume decreases. In the next few sections, the process of human breathing will be explained.

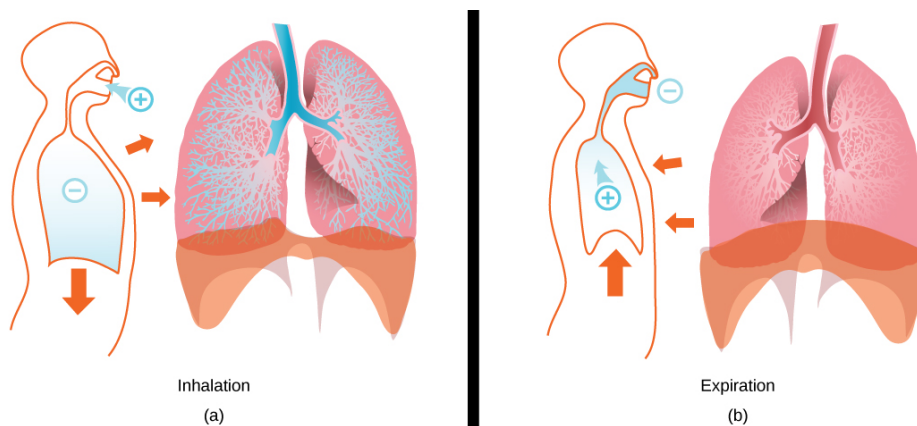
## The Mechanics of Human Breathing

Boyle's Law is the gas law that states that in a closed space, pressure and volume are inversely related. As volume decreases, pressure increases and vice versa ([\[link\]](#)). The relationship between gas pressure and volume helps to explain the mechanics of breathing.



This graph shows data from Boyle's original 1662 experiment, which shows that pressure and volume are inversely related. No units are given as Boyle used arbitrary units in his experiments.

There is always a slightly negative pressure within the thoracic cavity, which aids in keeping the airways of the lungs open. During inhalation, volume increases as a result of contraction of the diaphragm, and pressure decreases (according to Boyle's Law). This decrease of pressure in the thoracic cavity relative to the environment makes the cavity less than the atmosphere ([link](#)a). Because of this drop in pressure, air rushes into the respiratory passages. To increase the volume of the lungs, the chest wall expands. This results from the contraction of the **intercostal muscles**, the muscles that are connected to the rib cage. Lung volume expands because the diaphragm contracts and the intercostals muscles contract, thus expanding the thoracic cavity. This increase in the volume of the thoracic cavity lowers pressure compared to the atmosphere, so air rushes into the lungs, thus increasing its volume. The resulting increase in volume is largely attributed to an increase in alveolar space, because the bronchioles and bronchi are stiff structures that do not change in size.



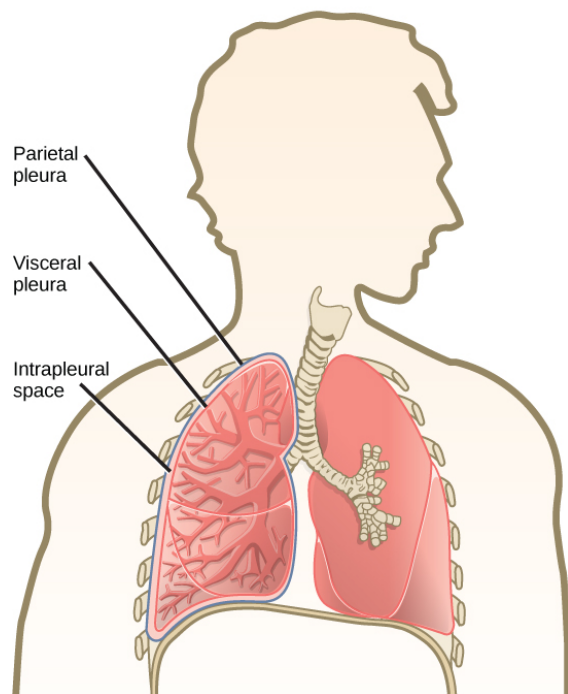
The lungs, chest wall, and diaphragm are all involved in respiration, both (a) inhalation and (b) expiration. (credit: modification of work by Mariana Ruiz Villareal)

The chest wall expands out and away from the lungs. The lungs are elastic; therefore, when air fills the lungs, the **elastic recoil** within the tissues of the lung exerts pressure back toward the interior of the lungs. These outward



and inward forces compete to inflate and deflate the lung with every breath. Upon exhalation, the lungs recoil to force the air out of the lungs, and the intercostal muscles relax, returning the chest wall back to its original position ([link](#)**b**). The diaphragm also relaxes and moves higher into the thoracic cavity. This increases the pressure within the thoracic cavity relative to the environment, and air rushes out of the lungs. The movement of air out of the lungs is a passive event. No muscles are contracting to expel the air.

Each lung is surrounded by an invaginated sac. The layer of tissue that covers the lung and dips into spaces is called the visceral **pleura**. A second layer of parietal pleura lines the interior of the thorax ([link](#)). The space between these layers, the **intrapleural space**, contains a small amount of fluid that protects the tissue and reduces the friction generated from rubbing the tissue layers together as the lungs contract and relax. **Pleurisy** results when these layers of tissue become inflamed; it is painful because the inflammation increases the pressure within the thoracic cavity and reduces the volume of the lung.



A tissue layer called pleura surrounds the lung and interior of the thoracic cavity.  
(credit: modification of work by NCI)

**Note:**

Link to Learning



View how Boyle's Law is related to breathing and watch a [video](#) on Boyle's Law.

[https://www.openstaxcollege.org/l/boyle\\_breathing](https://www.openstaxcollege.org/l/boyle_breathing)

## The Work of Breathing

The number of breaths per minute is the **respiratory rate**. On average, under non-exertion conditions, the human respiratory rate is 12–15 breaths/minute. The respiratory rate contributes to the **alveolar ventilation**, or how much air moves into and out of the alveoli. Alveolar ventilation prevents carbon dioxide buildup in the alveoli. There are two ways to keep the alveolar ventilation constant: increase the respiratory rate while decreasing the tidal volume of air per breath (shallow breathing), or decrease the respiratory rate while increasing the tidal volume per breath. In either case, the ventilation remains the same, but the work done and type of

work needed are quite different. Both tidal volume and respiratory rate are closely regulated when oxygen demand increases.

There are two types of work conducted during respiration, flow-resistive and elastic work. **Flow-resistive** refers to the work of the alveoli and tissues in the lung, whereas **elastic work** refers to the work of the intercostal muscles, chest wall, and diaphragm. Increasing the respiration rate increases the flow-resistive work of the airways and decreases the elastic work of the muscles. Decreasing the respiratory rate reverses the type of work required.

## Surfactant

The air-tissue/water interface of the alveoli has a high surface tension. This surface tension is similar to the surface tension of water at the liquid-air interface of a water droplet that results in the bonding of the water molecules together. **Surfactant** is a complex mixture of phospholipids and lipoproteins that works to reduce the surface tension that exists between the alveoli tissue and the air found within the alveoli. By lowering the surface tension of the alveolar fluid, it reduces the tendency of alveoli to collapse.

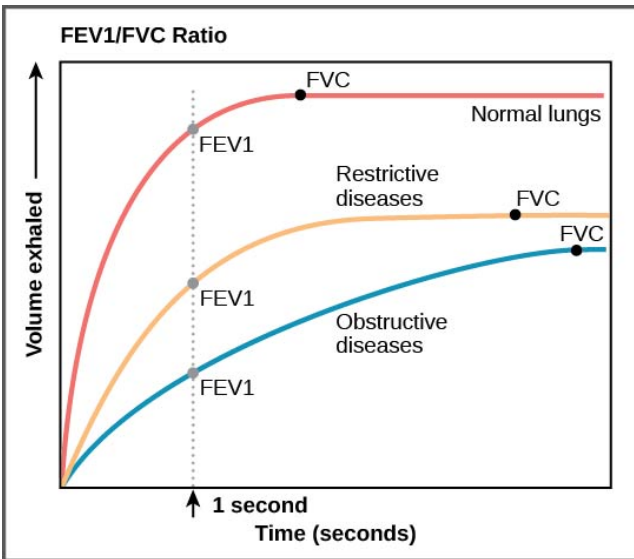
Surfactant works like a detergent to reduce the surface tension and allows for easier inflation of the airways. When a balloon is first inflated, it takes a large amount of effort to stretch the plastic and start to inflate the balloon. If a little bit of detergent was applied to the interior of the balloon, then the amount of effort or work needed to begin to inflate the balloon would decrease, and it would become much easier to start blowing up the balloon. This same principle applies to the airways. A small amount of surfactant to the airway tissues reduces the effort or work needed to inflate those airways. Babies born prematurely sometimes do not produce enough surfactant. As a result, they suffer from **respiratory distress syndrome**, because it requires more effort to inflate their lungs. Surfactant is also important for preventing collapse of small alveoli relative to large alveoli.

## Lung Resistance and Compliance

Pulmonary diseases reduce the rate of gas exchange into and out of the lungs. Two main causes of decreased gas exchange are **compliance** (how elastic the lung is) and **resistance** (how much obstruction exists in the airways). A change in either can dramatically alter breathing and the ability to take in oxygen and release carbon dioxide.

Examples of **restrictive diseases** are respiratory distress syndrome and pulmonary fibrosis. In both diseases, the airways are less compliant and they are stiff or fibrotic. There is a decrease in compliance because the lung tissue cannot bend and move. In these types of restrictive diseases, the intrapleural pressure is more positive and the airways collapse upon exhalation, which traps air in the lungs. Forced or **functional vital capacity (FVC)**, which is the amount of air that can be forcibly exhaled after taking the deepest breath possible, is much lower than in normal patients, and the time it takes to exhale most of the air is greatly prolonged ([\[link\]](#)). A patient suffering from these diseases cannot exhale the normal amount of air.

**Obstructive diseases** and conditions include emphysema, asthma, and pulmonary edema. In emphysema, which mostly arises from smoking tobacco, the walls of the alveoli are destroyed, decreasing the surface area for gas exchange. The overall compliance of the lungs is increased, because as the alveolar walls are damaged, lung elastic recoil decreases due to a loss of elastic fibers, and more air is trapped in the lungs at the end of exhalation. Asthma is a disease in which inflammation is triggered by environmental factors. Inflammation obstructs the airways. The obstruction may be due to edema (fluid accumulation), smooth muscle spasms in the walls of the bronchioles, increased mucus secretion, damage to the epithelia of the airways, or a combination of these events. Those with asthma or edema experience increased occlusion from increased inflammation of the airways. This tends to block the airways, preventing the proper movement of gases ([\[link\]](#)). Those with obstructive diseases have large volumes of air trapped after exhalation and breathe at a very high lung volume to compensate for the lack of airway recruitment.



The ratio of FEV1 (the amount of air that can be forcibly exhaled in one second after taking a deep breath) to FVC (the total amount of air that can be forcibly exhaled) can be used to diagnose whether a person has restrictive or obstructive lung disease. In restrictive lung disease, FVC is reduced but airways are not obstructed, so the person is able to expel air reasonably fast. In obstructive lung disease, airway obstruction results in slow exhalation as well as reduced FVC. Thus, the FEV1/FVC ratio is lower in persons with obstructive lung disease (less than 69 percent) than in persons with restrictive disease (88 to 90 percent).

## Dead Space: V/Q Mismatch

Pulmonary circulation pressure is very low compared to that of the systemic circulation. It is also independent of cardiac output. This is because of a phenomenon called **recruitment**, which is the process of opening airways that normally remain closed when cardiac output increases. As cardiac output increases, the number of capillaries and arteries that are perfused (filled with blood) increases. These capillaries and arteries are not always in use but are ready if needed. At times, however, there is a mismatch between the amount of air (ventilation, V) and the amount of blood (perfusion, Q) in the lungs. This is referred to as **ventilation/perfusion (V/Q) mismatch**.

There are two types of V/Q mismatch. Both produce **dead space**, regions of broken down or blocked lung tissue. Dead spaces can severely impact breathing, because they reduce the surface area available for gas diffusion. As a result, the amount of oxygen in the blood decreases, whereas the carbon dioxide level increases. Dead space is created when no ventilation and/or perfusion takes place. **Anatomical dead space** or anatomical shunt, arises from an anatomical failure, while **physiological dead space** or physiological shunt, arises from a functional impairment of the lung or arteries.

An example of an anatomical shunt is the effect of gravity on the lungs. The lung is particularly susceptible to changes in the magnitude and direction of gravitational forces. When someone is standing or sitting upright, the pleural pressure gradient leads to increased ventilation further down in the lung. As a result, the intrapleural pressure is more negative at the base of the lung than at the top, and more air fills the bottom of the lung than the top. Likewise, it takes less energy to pump blood to the bottom of the lung than to the top when in a prone position. Perfusion of the lung is not uniform while standing or sitting. This is a result of hydrostatic forces combined with the effect of airway pressure. An anatomical shunt develops because the ventilation of the airways does not match the perfusion of the arteries surrounding those airways. As a result, the rate of gas exchange is reduced. Note that this does not occur when lying down, because in this position, gravity does not preferentially pull the bottom of the lung down.

A physiological shunt can develop if there is infection or edema in the lung that obstructs an area. This will decrease ventilation but not affect perfusion; therefore, the V/Q ratio changes and gas exchange is affected.

The lung can compensate for these mismatches in ventilation and perfusion. If ventilation is greater than perfusion, the arterioles dilate and the bronchioles constrict. This increases perfusion and reduces ventilation. Likewise, if ventilation is less than perfusion, the arterioles constrict and the bronchioles dilate to correct the imbalance.

**Note:**

Link to Learning



View the mechanics of breathing.

<https://www.openstaxcollege.org/l/breathing>

## Section Summary

The structure of the lungs and thoracic cavity control the mechanics of breathing. Upon inspiration, the diaphragm contracts and lowers. The intercostal muscles contract and expand the chest wall outward. The intrapleural pressure drops, the lungs expand, and air is drawn into the airways. When exhaling, the intercostal muscles and diaphragm relax, returning the intrapleural pressure back to the resting state. The lungs recoil and airways close. The air passively exits the lung. There is high surface tension at the air-airway interface in the lung. Surfactant, a mixture of phospholipids and lipoproteins, acts like a detergent in the airways to reduce surface tension and allow for opening of the alveoli.

Breathing and gas exchange are both altered by changes in the compliance and resistance of the lung. If the compliance of the lung decreases, as occurs in restrictive diseases like fibrosis, the airways stiffen and collapse upon exhalation. Air becomes trapped in the lungs, making breathing more difficult. If resistance increases, as happens with asthma or emphysema, the airways become obstructed, trapping air in the lungs and causing breathing to become difficult. Alterations in the ventilation of the airways or perfusion of the arteries can affect gas exchange. These changes in ventilation and perfusion, called V/Q mismatch, can arise from anatomical or physiological changes.

## Review Questions

### Exercise:

**Problem:**How would paralysis of the diaphragm alter inspiration?

- a. It would prevent contraction of the intercostal muscles.
- b. It would prevent inhalation because the intrapleural pressure would not change.
- c. It would decrease the intrapleural pressure and allow more air to enter the lungs.
- d. It would slow expiration because the lung would not relax.

---

### Solution:

B

### Exercise:

**Problem:**Restrictive airway diseases \_\_\_\_\_.

- a. increase the compliance of the lung
- b. decrease the compliance of the lung
- c. increase the lung volume
- d. decrease the work of breathing



---

**Solution:**

B

**Exercise:**

**Problem:** Alveolar ventilation remains constant when \_\_\_\_\_.

- a. the respiratory rate is increased while the volume of air per breath is decreased
- b. the respiratory rate and the volume of air per breath are increased
- c. the respiratory rate is decreased while increasing the volume per breath
- d. both a and c

---

**Solution:**

D

**Free Response****Exercise:****Problem:**

How would increased airway resistance affect intrapleural pressure during inhalation?

---

**Solution:**

Increased airway resistance increases the volume and pressure in the lung; therefore, the intrapleural pressure would be less negative and breathing would be more difficult.

**Exercise:**

**Problem:**

Explain how a puncture to the thoracic cavity (from a knife wound, for instance) could alter the ability to inhale.

---

**Solution:**

A puncture to the thoracic cavity would equalize the pressure inside the thoracic cavity to the outside environment. For the lung to function properly, the intrapleural pressure must be negative. This is caused by the contraction of the diaphragm pulling the lungs down and drawing air into the lungs.

**Exercise:****Problem:**

When someone is standing, gravity stretches the bottom of the lung down toward the floor to a greater extent than the top of the lung. What implication could this have on the flow of air in the lungs? Where does gas exchange occur in the lungs?

---

**Solution:**

The lung is particularly susceptible to changes in the magnitude and direction of gravitational forces. When someone is standing or sitting upright, the pleural pressure gradient leads to increased ventilation further down in the lung.

**Glossary**

alveolar ventilation

how much air is in the alveoli

anatomical dead space

(also, anatomical shunt) region of the lung that lacks proper ventilation/perfusion due to an anatomical block

compliance

measurement of the elasticity of the lung

dead space

area in the lung that lacks proper ventilation or perfusion

elastic recoil

property of the lung that drives the lung tissue inward

elastic work

work conducted by the intercostal muscles, chest wall, and diaphragm

flow-resistive

work of breathing performed by the alveoli and tissues in the lung

functional vital capacity (FVC)

amount of air that can be forcibly exhaled after taking the deepest breath possible

intercostal muscle

muscle connected to the rib cage that contracts upon inspiration

intrapleural space

space between the layers of pleura

obstructive disease

disease (such as emphysema and asthma) that arises from obstruction of the airways; compliance increases in these diseases

physiological dead space

(also, physiological shunt) region of the lung that lacks proper ventilation/perfusion due to a physiological change in the lung (like inflammation or edema)

pleura

tissue layer that surrounds the lungs and lines the interior of the thoracic cavity

pleurisy

painful inflammation of the pleural tissue layers

recruitment

process of opening airways that normally remain closed when the cardiac output increases

resistance

measurement of lung obstruction

respiratory distress syndrome

disease that arises from a deficient amount of surfactant

respiratory rate

number of breaths per minute

restrictive disease

disease that results from a restriction and decreased compliance of the alveoli; respiratory distress syndrome and pulmonary fibrosis are examples

surfactant

detergent-like liquid in the airways that lowers the surface tension of the alveoli to allow for expansion

ventilation/perfusion (V/Q) mismatch

region of the lung that lacks proper alveolar ventilation (V) and/or arterial perfusion (Q)

## Transport of Gases in Human Bodily Fluids

By the end of this section, you will be able to:

- Describe how oxygen is bound to hemoglobin and transported to body tissues
- Explain how carbon dioxide is transported from body tissues to the lungs

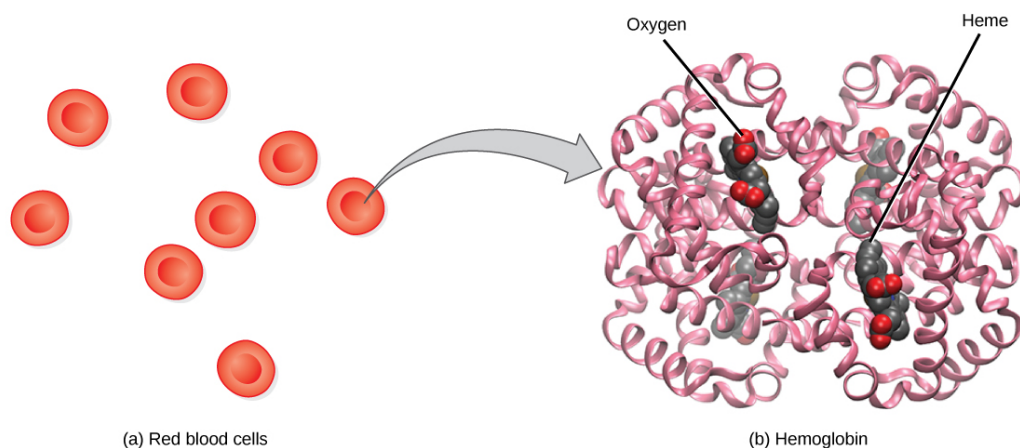
Once the oxygen diffuses across the alveoli, it enters the bloodstream and is transported to the tissues where it is unloaded, and carbon dioxide diffuses out of the blood and into the alveoli to be expelled from the body. Although gas exchange is a continuous process, the oxygen and carbon dioxide are transported by different mechanisms.

## Transport of Oxygen in the Blood

Although oxygen dissolves in blood, only a small amount of oxygen is transported this way. Only 1.5 percent of oxygen in the blood is dissolved directly into the blood itself. Most oxygen—98.5 percent—is bound to a protein called hemoglobin and carried to the tissues.

## Hemoglobin

**Hemoglobin**, or Hb, is a protein molecule found in red blood cells (erythrocytes) made of four subunits: two alpha subunits and two beta subunits ([\[link\]](#)). Each subunit surrounds a central **heme group** that contains iron and binds one oxygen molecule, allowing each hemoglobin molecule to bind four oxygen molecules. Molecules with more oxygen bound to the heme groups are brighter red. As a result, oxygenated arterial blood where the Hb is carrying four oxygen molecules is bright red, while venous blood that is deoxygenated is darker red.

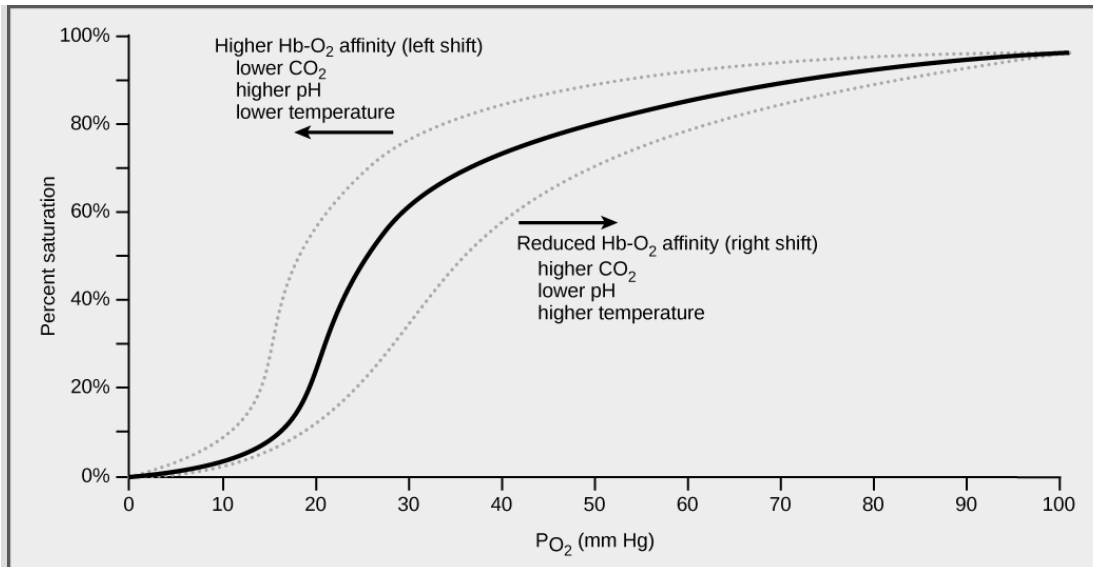


The protein inside (a) red blood cells that carries oxygen to cells and carbon dioxide to the lungs is (b) hemoglobin. Hemoglobin is made up of four symmetrical subunits and four heme groups. Iron associated with the heme binds oxygen. It is the iron in hemoglobin that gives blood its red color.

It is easier to bind a second and third oxygen molecule to Hb than the first molecule. This is because the hemoglobin molecule changes its shape, or conformation, as oxygen binds. The fourth oxygen is then more difficult to bind. The binding of oxygen to hemoglobin can be plotted as a function of the partial pressure of oxygen in the blood (x-axis) versus the relative Hb-oxygen saturation (y-axis). The resulting graph—an **oxygen dissociation curve**—is sigmoidal, or S-shaped ([\[link\]](#)). As the partial pressure of oxygen increases, the hemoglobin becomes increasingly saturated with oxygen.

### Note:

Art Connection



The oxygen dissociation curve demonstrates that, as the partial pressure of oxygen increases, more oxygen binds hemoglobin. However, the affinity of hemoglobin for oxygen may shift to the left or the right depending on environmental conditions.

The kidneys are responsible for removing excess  $H^+$  ions from the blood. If the kidneys fail, what would happen to blood pH and to hemoglobin affinity for oxygen?

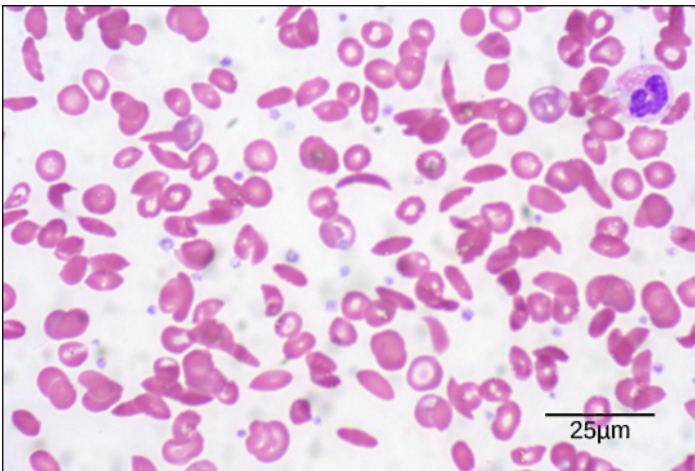
## Factors That Affect Oxygen Binding

The **oxygen-carrying capacity** of hemoglobin determines how much oxygen is carried in the blood. In addition to  $P_{O_2}$ , other environmental factors and diseases can affect oxygen carrying capacity and delivery.

Carbon dioxide levels, blood pH, and body temperature affect oxygen-carrying capacity ([\[link\]](#)). When carbon dioxide is in the blood, it reacts with water to form bicarbonate ( $HCO_3^-$ ) and hydrogen ions ( $H^+$ ). As the level of carbon dioxide in the blood increases, more  $H^+$  is produced and the pH decreases. This increase in carbon dioxide and subsequent decrease in pH reduce the affinity of hemoglobin for oxygen. The oxygen dissociates

from the Hb molecule, shifting the oxygen dissociation curve to the right. Therefore, more oxygen is needed to reach the same hemoglobin saturation level as when the pH was higher. A similar shift in the curve also results from an increase in body temperature. Increased temperature, such as from increased activity of skeletal muscle, causes the affinity of hemoglobin for oxygen to be reduced.

Diseases like sickle cell anemia and thalassemia decrease the blood's ability to deliver oxygen to tissues and its oxygen-carrying capacity. In **sickle cell anemia**, the shape of the red blood cell is crescent-shaped, elongated, and stiffened, reducing its ability to deliver oxygen ([\[link\]](#)). In this form, red blood cells cannot pass through the capillaries. This is painful when it occurs. **Thalassemia** is a rare genetic disease caused by a defect in either the alpha or the beta subunit of Hb. Patients with thalassemia produce a high number of red blood cells, but these cells have lower-than-normal levels of hemoglobin. Therefore, the oxygen-carrying capacity is diminished.



Individuals with sickle cell anemia have crescent-shaped red blood cells.  
(credit: modification of work by Ed Uthman; scale-bar data from Matt Russell)

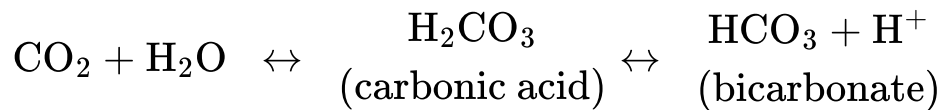


## Transport of Carbon Dioxide in the Blood

Carbon dioxide molecules are transported in the blood from body tissues to the lungs by one of three methods: dissolution directly into the blood, binding to hemoglobin, or carried as a bicarbonate ion. Several properties of carbon dioxide in the blood affect its transport. First, carbon dioxide is more soluble in blood than oxygen. About 5 to 7 percent of all carbon dioxide is dissolved in the plasma. Second, carbon dioxide can bind to plasma proteins or can enter red blood cells and bind to hemoglobin. This form transports about 10 percent of the carbon dioxide. When carbon dioxide binds to hemoglobin, a molecule called **carbaminohemoglobin** is formed. Binding of carbon dioxide to hemoglobin is reversible. Therefore, when it reaches the lungs, the carbon dioxide can freely dissociate from the hemoglobin and be expelled from the body.

Third, the majority of carbon dioxide molecules (85 percent) are carried as part of the **bicarbonate buffer system**. In this system, carbon dioxide diffuses into the red blood cells. **Carbonic anhydrase (CA)** within the red blood cells quickly converts the carbon dioxide into carbonic acid ( $\text{H}_2\text{CO}_3$ ). Carbonic acid is an unstable intermediate molecule that immediately dissociates into **bicarbonate ions** ( $\text{HCO}_3^-$ ) and hydrogen ( $\text{H}^+$ ) ions. Since carbon dioxide is quickly converted into bicarbonate ions, this reaction allows for the continued uptake of carbon dioxide into the blood down its concentration gradient. It also results in the production of  $\text{H}^+$  ions. If too much  $\text{H}^+$  is produced, it can alter blood pH. However, hemoglobin binds to the free  $\text{H}^+$  ions and thus limits shifts in pH. The newly synthesized bicarbonate ion is transported out of the red blood cell into the liquid component of the blood in exchange for a chloride ion ( $\text{Cl}^-$ ); this is called the **chloride shift**. When the blood reaches the lungs, the bicarbonate ion is transported back into the red blood cell in exchange for the chloride ion. The  $\text{H}^+$  ion dissociates from the hemoglobin and binds to the bicarbonate ion. This produces the carbonic acid intermediate, which is converted back into carbon dioxide through the enzymatic action of CA. The carbon dioxide produced is expelled through the lungs during exhalation.

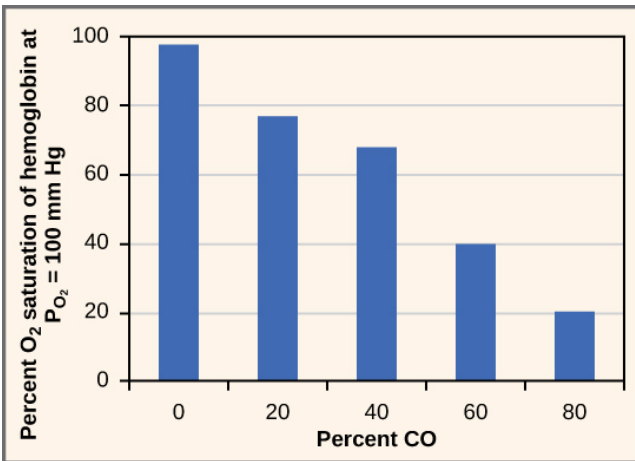
**Equation:**



The benefit of the bicarbonate buffer system is that carbon dioxide is “soaked up” into the blood with little change to the pH of the system. This is important because it takes only a small change in the overall pH of the body for severe injury or death to result. The presence of this bicarbonate buffer system also allows for people to travel and live at high altitudes: When the partial pressure of oxygen and carbon dioxide change at high altitudes, the bicarbonate buffer system adjusts to regulate carbon dioxide while maintaining the correct pH in the body.

## Carbon Monoxide Poisoning

While carbon dioxide can readily associate and dissociate from hemoglobin, other molecules such as carbon monoxide (CO) cannot. Carbon monoxide has a greater affinity for hemoglobin than oxygen. Therefore, when carbon monoxide is present, it binds to hemoglobin preferentially over oxygen. As a result, oxygen cannot bind to hemoglobin, so very little oxygen is transported through the body ([link](#)). Carbon monoxide is a colorless, odorless gas and is therefore difficult to detect. It is produced by gas-powered vehicles and tools. Carbon monoxide can cause headaches, confusion, and nausea; long-term exposure can cause brain damage or death. Administering 100 percent (pure) oxygen is the usual treatment for carbon monoxide poisoning. Administration of pure oxygen speeds up the separation of carbon monoxide from hemoglobin.



As percent CO increases, the oxygen saturation of hemoglobin decreases.

## Section Summary

Hemoglobin is a protein found in red blood cells that is comprised of two alpha and two beta subunits that surround an iron-containing heme group. Oxygen readily binds this heme group. The ability of oxygen to bind increases as more oxygen molecules are bound to heme. Disease states and altered conditions in the body can affect the binding ability of oxygen, and increase or decrease its ability to dissociate from hemoglobin.

Carbon dioxide can be transported through the blood via three methods. It is dissolved directly in the blood, bound to plasma proteins or hemoglobin, or converted into bicarbonate. The majority of carbon dioxide is transported as part of the bicarbonate system. Carbon dioxide diffuses into red blood cells. Inside, carbonic anhydrase converts carbon dioxide into carbonic acid ( $\text{H}_2\text{CO}_3$ ), which is subsequently hydrolyzed into bicarbonate ( $\text{HCO}_3^-$ ) and  $\text{H}^+$ . The  $\text{H}^+$  ion binds to hemoglobin in red blood cells, and bicarbonate is transported out of the red blood cells in exchange for a chloride ion. This is called the chloride shift. Bicarbonate leaves the red blood cells and enters the blood plasma. In the lungs, bicarbonate is transported back into the red

blood cells in exchange for chloride. The  $H^+$  dissociates from hemoglobin and combines with bicarbonate to form carbonic acid with the help of carbonic anhydrase, which further catalyzes the reaction to convert carbonic acid back into carbon dioxide and water. The carbon dioxide is then expelled from the lungs.

## Art Connections

### Exercise:

#### Problem:

[\[link\]](#) The kidneys are responsible for removing excess  $H^+$  ions from the blood. If the kidneys fail, what would happen to blood pH and to hemoglobin affinity for oxygen?

---

#### Solution:

[\[link\]](#) The blood pH will drop and hemoglobin affinity for oxygen will decrease.

## Review Questions

### Exercise:

#### Problem:

Which of the following will NOT facilitate the transfer of oxygen to tissues?

- a. decreased body temperature
  - b. decreased pH of the blood
  - c. increased carbon dioxide
  - d. increased exercise
- 

#### Solution:

A

**Exercise:**

**Problem:**

The majority of carbon dioxide in the blood is transported by \_\_\_\_\_.

- a. binding to hemoglobin
- b. dissolution in the blood
- c. conversion to bicarbonate
- d. binding to plasma proteins

---

**Solution:**

C

**Exercise:**

**Problem:**

The majority of oxygen in the blood is transported by \_\_\_\_\_.

- a. dissolution in the blood
- b. being carried as bicarbonate ions
- c. binding to blood plasma
- d. binding to hemoglobin

---

**Solution:**

D

**Free Response**

**Exercise:**

**Problem:**

What would happen if no carbonic anhydrase were present in red blood cells?

---

**Solution:**

Without carbonic anhydrase, carbon dioxide would not be hydrolyzed into carbonic acid or bicarbonate. Therefore, very little carbon dioxide (only 15 percent) would be transported in the blood away from the tissues.

**Exercise:****Problem:**

How does the administration of 100 percent oxygen save a patient from carbon monoxide poisoning? Why wouldn't giving carbon dioxide work?

---

**Solution:**

Carbon monoxide has a higher affinity for hemoglobin than oxygen. This means that carbon monoxide will preferentially bind to hemoglobin over oxygen. Administration of 100 percent oxygen is an effective therapy because at that concentration, oxygen will displace the carbon monoxide from the hemoglobin.

**Glossary**

bicarbonate buffer system

system in the blood that absorbs carbon dioxide and regulates pH levels

bicarbonate ( $\text{HCO}_3^-$ ) ion

ion created when carbonic acid dissociates into  $\text{H}^+$  and ( $\text{HCO}_3^-$ )

carbaminohemoglobin

molecule that forms when carbon dioxide binds to hemoglobin

carbonic anhydrase (CA)

enzyme that catalyzes carbon dioxide and water into carbonic acid

chloride shift

chloride shift exchange of chloride for bicarbonate into or out of the red blood cell

heme group

centralized iron-containing group that is surrounded by the alpha and beta subunits of hemoglobin

hemoglobin

molecule in red blood cells that can bind oxygen, carbon dioxide, and carbon monoxide

oxygen-carrying capacity

amount of oxygen that can be transported in the blood

oxygen dissociation curve

curve depicting the affinity of oxygen for hemoglobin

sickle cell anemia

genetic disorder that affects the shape of red blood cells, and their ability to transport oxygen and move through capillaries

thalassemia

rare genetic disorder that results in mutation of the alpha or beta subunits of hemoglobin, creating smaller red blood cells with less hemoglobin

## Introduction

class="introduction"

Just as  
highway  
systems  
transport  
people and  
goods  
through a  
complex  
network, the  
circulatory  
system  
transports  
nutrients,  
gases, and  
wastes  
throughout  
the animal  
body.

(credit:  
modificatio  
n of work  
by Andrey  
Belenko)





Most animals are complex multicellular organisms that require a mechanism for transporting nutrients throughout their bodies and removing waste products. The circulatory system has evolved over time from simple diffusion through cells in the early evolution of animals to a complex network of blood vessels that reach all parts of the human body. This extensive network supplies the cells, tissues, and organs with oxygen and nutrients, and removes carbon dioxide and waste, which are byproducts of respiration.

At the core of the human circulatory system is the heart. The size of a clenched fist, the human heart is protected beneath the rib cage. Made of specialized and unique cardiac muscle, it pumps blood throughout the body and to the heart itself. Heart contractions are driven by intrinsic electrical impulses that the brain and endocrine hormones help to regulate. Understanding the heart's basic anatomy and function is important to understanding the body's circulatory and respiratory systems.

Gas exchange is one essential function of the circulatory system. A circulatory system is not needed in organisms with no specialized respiratory organs because oxygen and carbon dioxide diffuse directly between their body tissues and the external environment. However, in organisms that possess lungs and gills, oxygen must be transported from these specialized respiratory organs to the body tissues via a circulatory

system. Therefore, circulatory systems have had to evolve to accommodate the great diversity of body sizes and body types present among animals.

## Overview of the Circulatory System

By the end of this section, you will be able to:

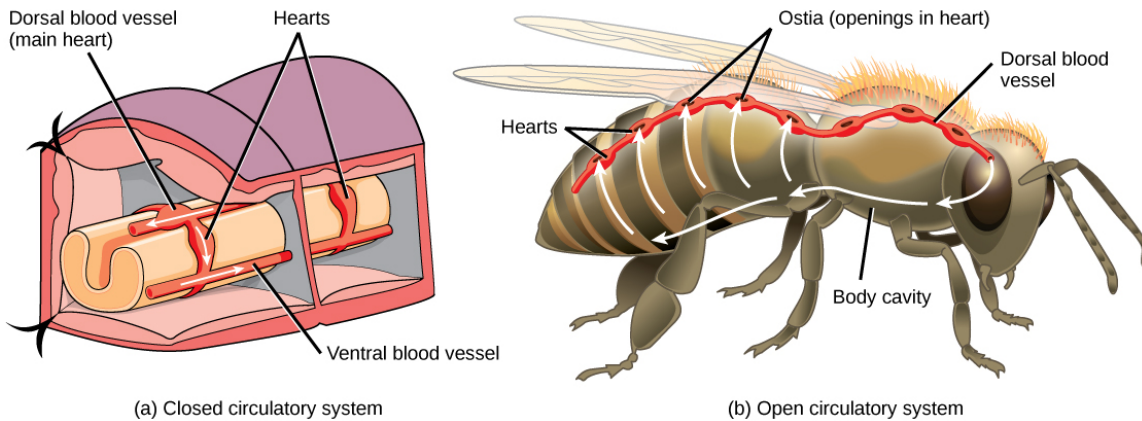
- Describe an open and closed circulatory system
- Describe interstitial fluid and hemolymph
- Compare and contrast the organization and evolution of the vertebrate circulatory system.

In all animals, except a few simple types, the circulatory system is used to transport nutrients and gases through the body. Simple diffusion allows some water, nutrient, waste, and gas exchange into primitive animals that are only a few cell layers thick; however, bulk flow is the only method by which the entire body of larger more complex organisms is accessed.

## Circulatory System Architecture

The circulatory system is effectively a network of cylindrical vessels: the arteries, veins, and capillaries that emanate from a pump, the heart. In all vertebrate organisms, as well as some invertebrates, this is a closed-loop system, in which the blood is not free in a cavity. In a **closed circulatory system**, blood is contained inside blood vessels and circulates **unidirectionally** from the heart around the systemic circulatory route, then returns to the heart again, as illustrated in [\[link\]](#)**a**. As opposed to a closed system, arthropods—including insects, crustaceans, and most mollusks—have an open circulatory system, as illustrated in [\[link\]](#)**b**. In an **open circulatory system**, the blood is not enclosed in the blood vessels but is pumped into a cavity called a **hemocoel** and is called **hemolymph** because the blood mixes with the **interstitial fluid**. As the heart beats and the animal moves, the hemolymph circulates around the organs within the body cavity and then reenters the hearts through openings called **ostia**. This movement allows for gas and nutrient exchange. An open circulatory system does not use as much energy as a closed system to operate or to maintain; however, there is a trade-off with the amount of blood that can be moved to metabolically active organs and tissues that require high levels of oxygen. In fact, one reason that insects with wing spans of up to two feet wide (70 cm) are not around today is probably because they were outcompeted by the arrival of birds 150 million years ago. Birds, having a

closed circulatory system, are thought to have moved more agilely, allowing them to get food faster and possibly to prey on the insects.



In (a) closed circulatory systems, the heart pumps blood through vessels that are separate from the interstitial fluid of the body.

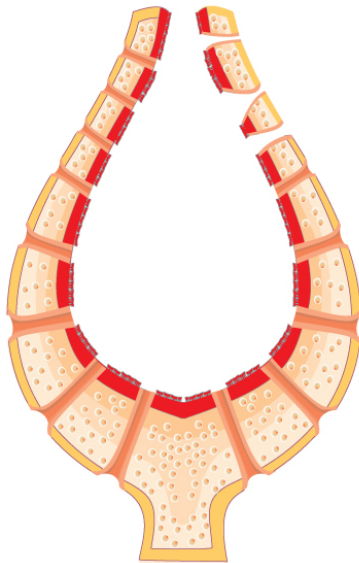
Most vertebrates and some invertebrates, like this annelid earthworm, have a closed circulatory system. In (b) open circulatory systems, a fluid called hemolymph is pumped through a blood vessel that empties into the body cavity. Hemolymph returns to the blood vessel through openings called ostia.

Arthropods like this bee and most mollusks have open circulatory systems.

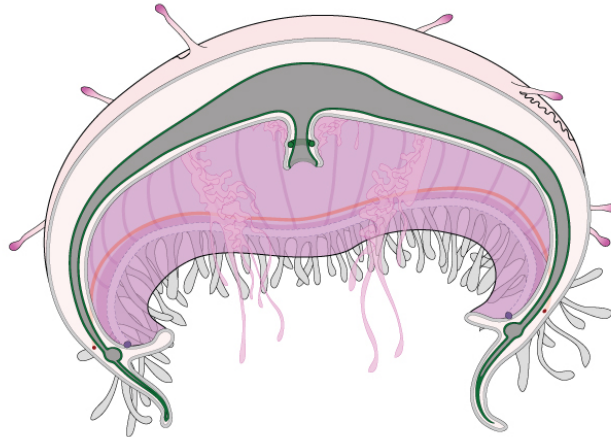
## Circulatory System Variation in Animals

The circulatory system varies from simple systems in invertebrates to more complex systems in vertebrates. The simplest animals, such as the sponges (Porifera) and rotifers (Rotifera), do not need a circulatory system because diffusion allows adequate exchange of water, nutrients, and waste, as well as dissolved gases, as shown in [link](#)a. Organisms that are more complex but still only have two layers of cells in their body plan, such as jellies (Cnidaria) and comb jellies (Ctenophora) also use diffusion through their epidermis and internally through the gastrovascular compartment. Both

their internal and external tissues are bathed in an aqueous environment and exchange fluids by diffusion on both sides, as illustrated in [\[link\]](#)**b**. Exchange of fluids is assisted by the pulsing of the jellyfish body.



(a) Sponge



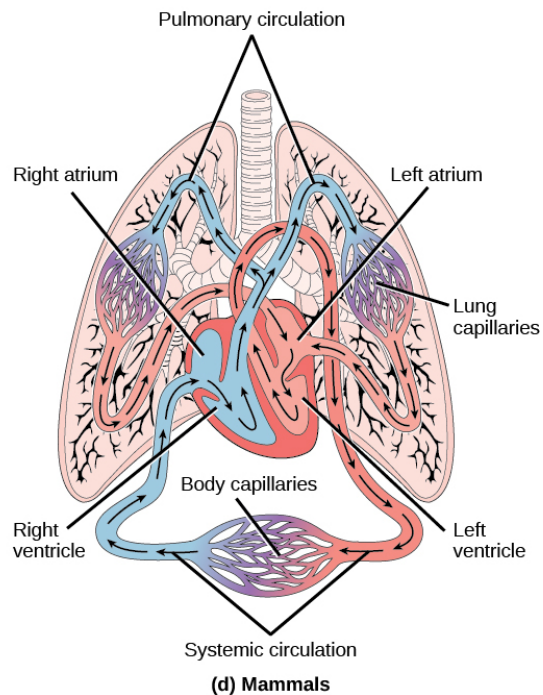
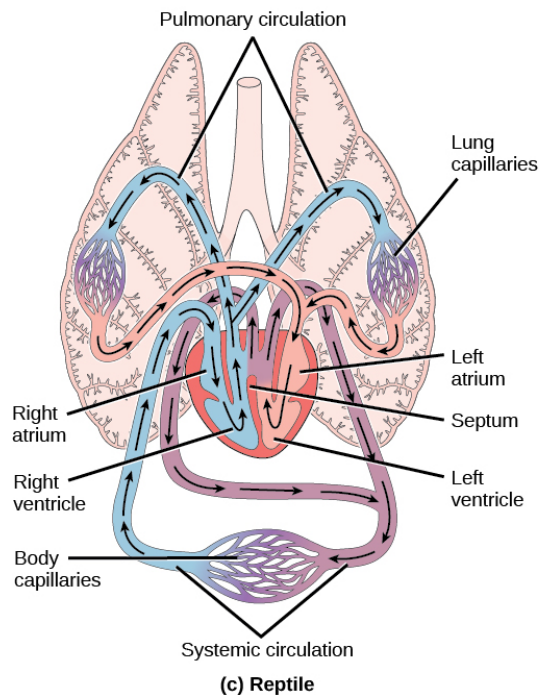
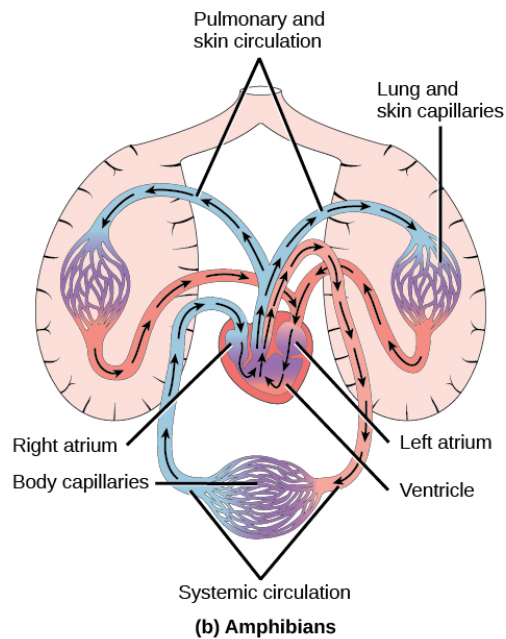
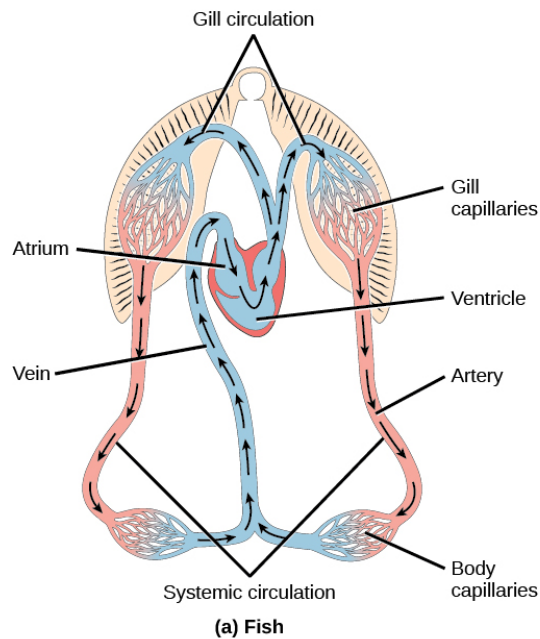
(b) Jellyfish

Simple animals consisting of a single cell layer such as the (a) sponge or only a few cell layers such as the (b) jellyfish do not have a circulatory system. Instead, gases, nutrients, and wastes are exchanged by diffusion.

For more complex organisms, diffusion is not efficient for cycling gases, nutrients, and waste effectively through the body; therefore, more complex circulatory systems evolved. Most arthropods and many mollusks have open circulatory systems. In an open system, an elongated beating heart pushes the hemolymph through the body and muscle contractions help to move fluids. The larger more complex crustaceans, including lobsters, have developed arterial-like vessels to push blood through their bodies, and the most active mollusks, such as squids, have evolved a closed circulatory system and are able to move rapidly to catch prey. Closed circulatory systems are a characteristic of vertebrates; however, there are significant



differences in the structure of the heart and the circulation of blood between the different vertebrate groups due to adaptation during evolution and associated differences in anatomy. [\[link\]](#) illustrates the basic circulatory systems of some vertebrates: fish, amphibians, reptiles, and mammals.



(a) Fish have the simplest circulatory systems of the vertebrates: blood flows unidirectionally from the two-chambered heart through the gills and then the rest of the body. (b) Amphibians have two circulatory routes: one for oxygenation of the blood through the lungs and skin, and the other to take oxygen to the rest of the body. The blood is pumped from a three-chambered heart with two atria and a single ventricle. (c) Reptiles also have two circulatory routes; however, blood is only oxygenated through the lungs. The heart is three chambered, but the ventricles are partially separated so some mixing of oxygenated and deoxygenated blood occurs except in crocodilians and birds. (d) Mammals and birds have the most efficient heart with four chambers that completely separate the oxygenated and deoxygenated blood; it pumps only oxygenated blood through the body and deoxygenated blood to the lungs.

As illustrated in [\[link\]](#) a Fish have a single circuit for blood flow and a two-chambered heart that has only a single atrium and a single ventricle. The atrium collects blood that has returned from the body and the ventricle pumps the blood to the gills where gas exchange occurs and the blood is re-oxygenated; this is called **gill circulation**. The blood then continues through the rest of the body before arriving back at the atrium; this is called **systemic circulation**. This unidirectional flow of blood produces a gradient of oxygenated to deoxygenated blood around the fish's systemic circuit. The result is a limit in the amount of oxygen that can reach some of the organs and tissues of the body, reducing the overall metabolic capacity of fish.

In amphibians, reptiles, birds, and mammals, blood flow is directed in two circuits: one through the lungs and back to the heart, which is called **pulmonary circulation**, and the other throughout the rest of the body and its organs including the brain (systemic circulation). In amphibians, gas exchange also occurs through the skin during pulmonary circulation and is referred to as **pulmocutaneous circulation**.

As shown in [\[link\]b](#), amphibians have a three-chambered heart that has two atria and one ventricle rather than the two-chambered heart of fish. The two **atria** (superior heart chambers) receive blood from the two different circuits (the lungs and the systems), and then there is some mixing of the blood in the heart's **ventricle** (inferior heart chamber), which reduces the efficiency of oxygenation. The advantage to this arrangement is that high pressure in the vessels pushes blood to the lungs and body. The mixing is mitigated by a ridge within the ventricle that diverts oxygen-rich blood through the systemic circulatory system and deoxygenated blood to the pulmocutaneous circuit. For this reason, amphibians are often described as having **double circulation**.

Most reptiles also have a three-chambered heart similar to the amphibian heart that directs blood to the pulmonary and systemic circuits, as shown in [\[link\]c](#). The ventricle is divided more effectively by a partial septum, which results in less mixing of oxygenated and deoxygenated blood. Some reptiles (alligators and crocodiles) are the most primitive animals to exhibit a four-chambered heart. Crocodilians have a unique circulatory mechanism where the heart shunts blood from the lungs toward the stomach and other organs during long periods of submergence, for instance, while the animal waits for prey or stays underwater waiting for prey to rot. One adaptation includes two main arteries that leave the same part of the heart: one takes blood to the lungs and the other provides an alternate route to the stomach and other parts of the body. Two other adaptations include a hole in the heart between the two ventricles, called the foramen of Panizza, which allows blood to move from one side of the heart to the other, and specialized connective tissue that slows the blood flow to the lungs. Together these adaptations have made crocodiles and alligators one of the most evolutionarily successful animal groups on earth.

In mammals and birds, the heart is also divided into four chambers: two atria and two ventricles, as illustrated in [\[link\]d](#). The oxygenated blood is separated from the deoxygenated blood, which improves the efficiency of double circulation and is probably required for the warm-blooded lifestyle of mammals and birds. The four-chambered heart of birds and mammals evolved independently from a three-chambered heart. The independent



evolution of the same or a similar biological trait is referred to as convergent evolution.

## Section Summary

In most animals, the circulatory system is used to transport blood through the body. Some primitive animals use diffusion for the exchange of water, nutrients, and gases. However, complex organisms use the circulatory system to carry gases, nutrients, and waste through the body. Circulatory systems may be open (mixed with the interstitial fluid) or closed (separated from the interstitial fluid). Closed circulatory systems are a characteristic of vertebrates; however, there are significant differences in the structure of the heart and the circulation of blood between the different vertebrate groups due to adaptations during evolution and associated differences in anatomy. Fish have a two-chambered heart with unidirectional circulation. Amphibians have a three-chambered heart, which has some mixing of the blood, and they have double circulation. Most non-avian reptiles have a three-chambered heart, but have little mixing of the blood; they have double circulation. Mammals and birds have a four-chambered heart with no mixing of the blood and double circulation.

## Review Questions

### Exercise:

#### Problem:

Why are open circulatory systems advantageous to some animals?

- a. They use less metabolic energy.
- b. They help the animal move faster.
- c. They do not need a heart.
- d. They help large insects develop.

---

#### Solution:

A

**Exercise:**

**Problem:**

Some animals use diffusion instead of a circulatory system. Examples include:

- a. birds and jellyfish
- b. flatworms and arthropods
- c. mollusks and jellyfish
- d. None of the above

---

**Solution:**

D

**Exercise:**

**Problem:**

Blood flow that is directed through the lungs and back to the heart is called \_\_\_\_\_.

- a. unidirectional circulation
- b. gill circulation
- c. pulmonary circulation
- d. pulmocutaneous circulation

---

**Solution:**

C

**Free Response**

**Exercise:**

**Problem:**Describe a closed circulatory system.

---

**Solution:**

A closed circulatory system is a closed-loop system, in which blood is not free in a cavity. Blood is separate from the bodily interstitial fluid and contained within blood vessels. In this type of system, blood circulates unidirectionally from the heart around the systemic circulatory route, and then returns to the heart.

**Exercise:**

**Problem:**Describe systemic circulation.

---

**Solution:**

Systemic circulation flows through the systems of the body. The blood flows away from the heart to the brain, liver, kidneys, stomach, and other organs, the limbs, and the muscles of the body; it then returns to the heart.

## Glossary

atrium

(plural: atria) chamber of the heart that receives blood from the veins and sends blood to the ventricles

closed circulatory system

system in which the blood is separated from the bodily interstitial fluid and contained in blood vessels

double circulation

flow of blood in two circuits: the pulmonary circuit through the lungs and the systemic circuit through the organs and body

gill circulation

circulatory system that is specific to animals with gills for gas exchange; the blood flows through the gills for oxygenation

hemocoel

cavity into which blood is pumped in an open circulatory system

hemolymph

mixture of blood and interstitial fluid that is found in insects and other arthropods as well as most mollusks

interstitial fluid

fluid between cells

open circulatory system

system in which the blood is mixed with interstitial fluid and directly covers the organs

ostium

(plural: ostia) holes between blood vessels that allow the movement of hemolymph through the body of insects, arthropods, and mollusks with open circulatory systems

pulmocutaneous circulation

circulatory system in amphibians; the flow of blood to the lungs and the moist skin for gas exchange

pulmonary circulation

flow of blood away from the heart through the lungs where oxygenation occurs and then returns to the heart again

systemic circulation

flow of blood away from the heart to the brain, liver, kidneys, stomach, and other organs, the limbs, and the muscles of the body, and then the return of this blood to the heart

unidirectional circulation

flow of blood in a single circuit; occurs in fish where the blood flows through the gills, then past the organs and the rest of the body, before

returning to the heart

ventricle

(heart) large inferior chamber of the heart that pumps blood into arteries

## Components of the Blood

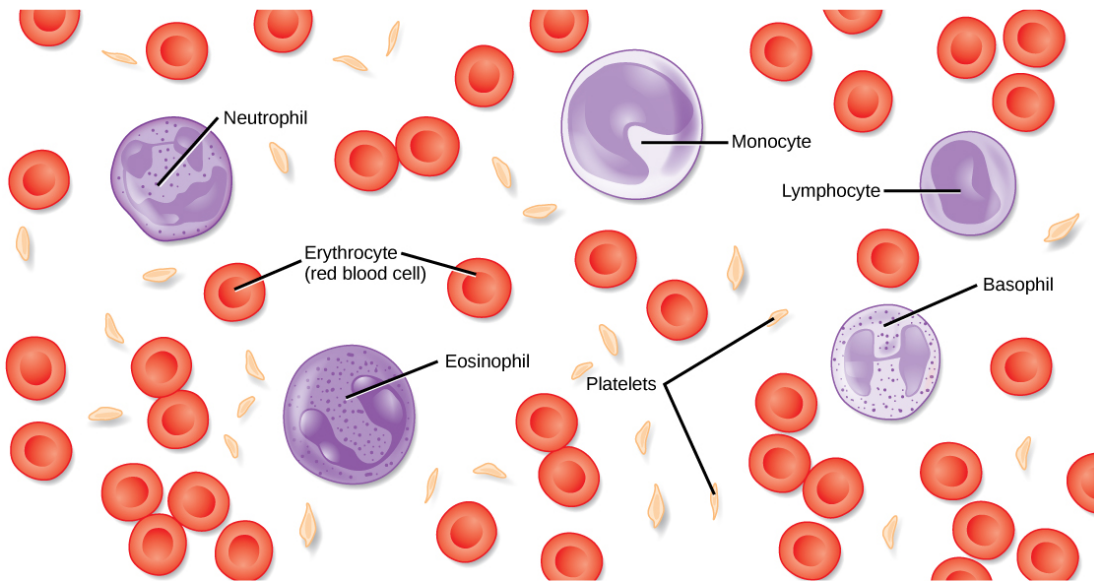
By the end of this section, you will be able to:

- List the basic components of the blood
- Compare red and white blood cells
- Describe blood plasma and serum

Hemoglobin is responsible for distributing oxygen, and to a lesser extent, carbon dioxide, throughout the circulatory systems of humans, vertebrates, and many invertebrates. The blood is more than the proteins, though. Blood is actually a term used to describe the liquid that moves through the vessels and includes **plasma** (the liquid portion, which contains water, proteins, salts, lipids, and glucose) and the cells (red and white cells) and cell fragments called **platelets**. Blood plasma is actually the dominant component of blood and contains the water, proteins, electrolytes, lipids, and glucose. The cells are responsible for carrying the gases (red cells) and immune the response (white). The platelets are responsible for blood clotting. Interstitial fluid that surrounds cells is separate from the blood, but in hemolymph, they are combined. In humans, cellular components make up approximately 45 percent of the blood and the liquid plasma 55 percent. Blood is 20 percent of a person's extracellular fluid and eight percent of weight.

## The Role of Blood in the Body

Blood, like the human blood illustrated in [\[link\]](#) is important for regulation of the body's systems and homeostasis. Blood helps maintain homeostasis by stabilizing pH, temperature, osmotic pressure, and by eliminating excess heat. Blood supports growth by distributing nutrients and hormones, and by removing waste. Blood plays a protective role by transporting clotting factors and platelets to prevent blood loss and transporting the disease-fighting agents or **white blood cells** to sites of infection.



The cells and cellular components of human blood are shown. Red blood cells deliver oxygen to the cells and remove carbon dioxide. White blood cells—including neutrophils, monocytes, lymphocytes, eosinophils, and basophils—are involved in the immune response. Platelets form clots that prevent blood loss after injury.

## Red Blood Cells

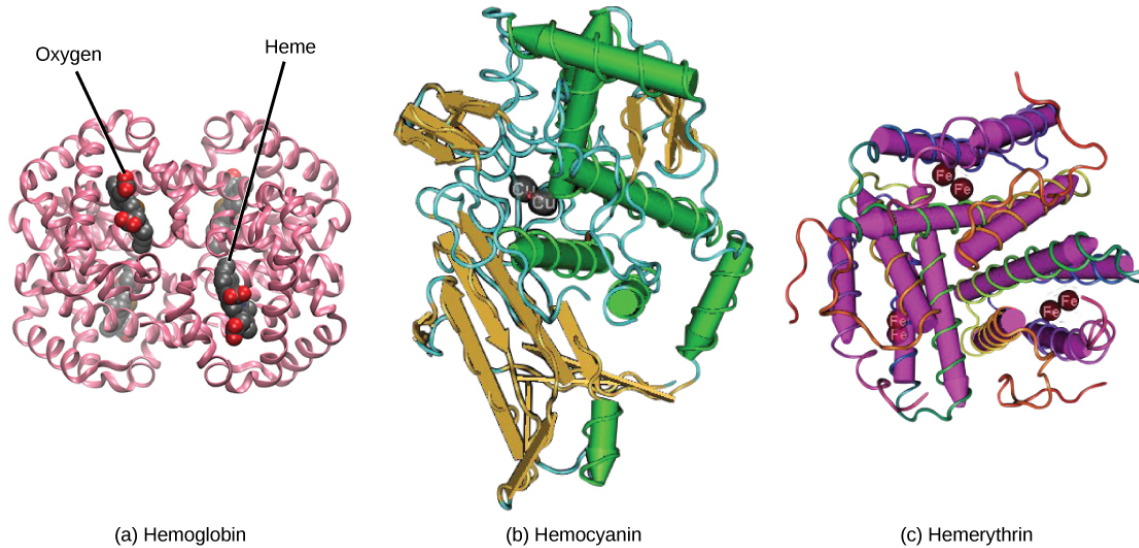
**Red blood cells**, or erythrocytes (erythro- = “red”; -cyte = “cell”), are specialized cells that circulate through the body delivering oxygen to cells; they are formed from stem cells in the bone marrow. In mammals, red blood cells are small biconcave cells that at maturity do not contain a nucleus or mitochondria and are only 7–8  $\mu\text{m}$  in size. In birds and non-avian reptiles, a nucleus is still maintained in red blood cells.

The red coloring of blood comes from the iron-containing protein hemoglobin, illustrated in [\[link\]](#)**a**. The principal job of this protein is to carry oxygen, but it also transports carbon dioxide as well. Hemoglobin is packed into red blood cells at a rate of about 250 million molecules of hemoglobin per cell. Each hemoglobin molecule binds four oxygen

molecules so that each red blood cell carries one billion molecules of oxygen. There are approximately 25 trillion red blood cells in the five liters of blood in the human body, which could carry up to 25 sextillion ( $25 \times 10^{21}$ ) molecules of oxygen in the body at any time. In mammals, the lack of organelles in erythrocytes leaves more room for the hemoglobin molecules, and the lack of mitochondria also prevents use of the oxygen for metabolic respiration. Only mammals have anucleated red blood cells, and some mammals (camels, for instance) even have nucleated red blood cells. The advantage of nucleated red blood cells is that these cells can undergo mitosis. Anucleated red blood cells metabolize anaerobically (without oxygen), making use of a primitive metabolic pathway to produce ATP and increase the efficiency of oxygen transport.

Not all organisms use hemoglobin as the method of oxygen transport. Invertebrates that utilize hemolymph rather than blood use different pigments to bind to the oxygen. These pigments use copper or iron to the oxygen. Invertebrates have a variety of other respiratory pigments. Hemocyanin, a blue-green, copper-containing protein, illustrated in [\[link\]](#)**b** is found in mollusks, crustaceans, and some of the arthropods. Chlorocruorin, a green-colored, iron-containing pigment is found in four families of polychaete tubeworms. Hemerythrin, a red, iron-containing protein is found in some polychaete worms and annelids and is illustrated in [\[link\]](#)**c**. Despite the name, hemerythrin does not contain a heme group and its oxygen-carrying capacity is poor compared to hemoglobin.





In most vertebrates, (a) hemoglobin delivers oxygen to the body and removes some carbon dioxide. Hemoglobin is composed of four protein subunits, two alpha chains and two beta chains, and a heme group that has iron associated with it. The iron reversibly associates with oxygen, and in so doing is oxidized from  $\text{Fe}^{2+}$  to  $\text{Fe}^{3+}$ . In most mollusks and some arthropods, (b) hemocyanin delivers oxygen. Unlike hemoglobin, hemolymph is not carried in blood cells, but floats free in the hemolymph. Copper instead of iron binds the oxygen, giving the hemolymph a blue-green color. In annelids, such as the earthworm, and some other invertebrates, (c) hemerythrin carries oxygen. Like hemoglobin, hemerythrin is carried in blood cells and has iron associated with it, but despite its name, hemerythrin does not contain heme.

The small size and large surface area of red blood cells allows for rapid diffusion of oxygen and carbon dioxide across the plasma membrane. In the lungs, carbon dioxide is released and oxygen is taken in by the blood. In the tissues, oxygen is released from the blood and carbon dioxide is bound for transport back to the lungs. Studies have found that hemoglobin also binds nitrous oxide (NO). NO is a vasodilator that relaxes the blood vessels and capillaries and may help with gas exchange and the passage of red blood cells through narrow vessels. Nitroglycerin, a heart medication for angina

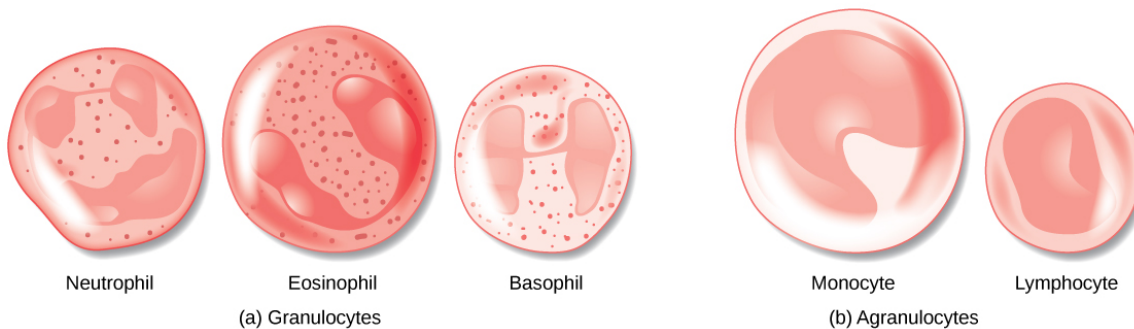
and heart attacks, is converted to NO to help relax the blood vessels and increase oxygen flow through the body.

A characteristic of red blood cells is their glycolipid and glycoprotein coating; these are lipids and proteins that have carbohydrate molecules attached. In humans, the surface glycoproteins and glycolipids on red blood cells vary between individuals, producing the different blood types, such as A, B, and O. Red blood cells have an average life span of 120 days, at which time they are broken down and recycled in the liver and spleen by phagocytic macrophages, a type of white blood cell.

## **White Blood Cells**

White blood cells, also called leukocytes (leuko = white), make up approximately one percent by volume of the cells in blood. The role of white blood cells is very different than that of red blood cells: they are primarily involved in the immune response to identify and target pathogens, such as invading bacteria, viruses, and other foreign organisms. White blood cells are formed continually; some only live for hours or days, but some live for years.

The morphology of white blood cells differs significantly from red blood cells. They have nuclei and do not contain hemoglobin. The different types of white blood cells are identified by their microscopic appearance after histologic staining, and each has a different specialized function. The two main groups, both illustrated in [\[link\]](#) are the granulocytes, which include the neutrophils, eosinophils, and basophils, and the agranulocytes, which include the monocytes and lymphocytes.



(a) Granulocytes—including neutrophils, eosinophils and basophils—are characterized by a lobed nucleus and granular inclusions in the cytoplasm. Granulocytes are typically first-responders during injury or infection. (b) Agranulocytes include lymphocytes and monocytes. Lymphocytes, including B and T cells, are responsible for adaptive immune response. Monocytes differentiate into macrophages and dendritic cells, which in turn respond to infection or injury.

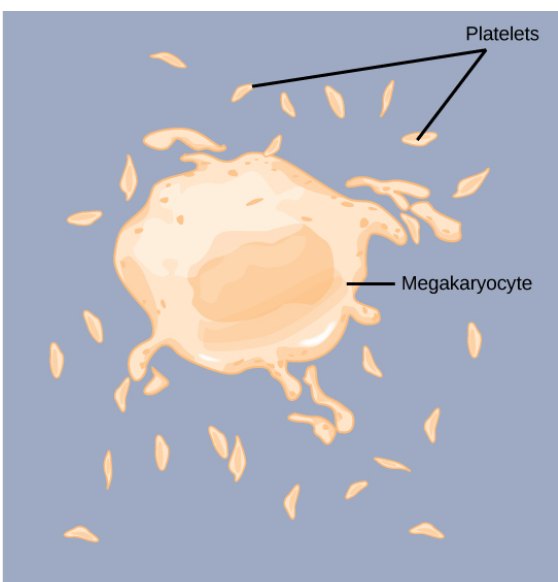
Granulocytes contain granules in their cytoplasm; the agranulocytes are so named because of the lack of granules in their cytoplasm. Some leukocytes become macrophages that either stay at the same site or move through the blood stream and gather at sites of infection or inflammation where they are attracted by chemical signals from foreign particles and damaged cells. Lymphocytes are the primary cells of the immune system and include B cells, T cells, and natural killer cells. B cells destroy bacteria and inactivate their toxins. They also produce antibodies. T cells attack viruses, fungi, some bacteria, transplanted cells, and cancer cells. T cells attack viruses by releasing toxins that kill the viruses. Natural killer cells attack a variety of infectious microbes and certain tumor cells.

One reason that HIV poses significant management challenges is because the virus directly targets T cells by gaining entry through a receptor. Once inside the cell, HIV then multiplies using the T cell's own genetic machinery. After the HIV virus replicates, it is transmitted directly from the infected T cell to macrophages. The presence of HIV can remain

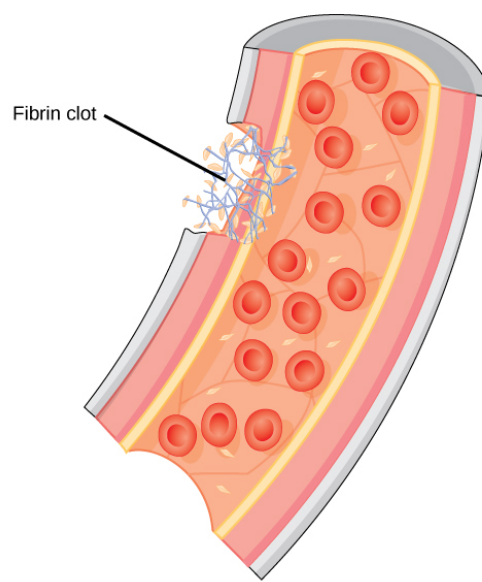
unrecognized for an extensive period of time before full disease symptoms develop.

## Platelets and Coagulation Factors

Blood must clot to heal wounds and prevent excess blood loss. Small cell fragments called platelets (thrombocytes) are attracted to the wound site where they adhere by extending many projections and releasing their contents. These contents activate other platelets and also interact with other coagulation factors, which convert fibrinogen, a water-soluble protein present in blood serum into fibrin (a non-water soluble protein), causing the blood to clot. Many of the clotting factors require vitamin K to work, and vitamin K deficiency can lead to problems with blood clotting. Many platelets converge and stick together at the wound site forming a platelet plug (also called a fibrin clot), as illustrated in [\[link\]](#)b. The plug or clot lasts for a number of days and stops the loss of blood. Platelets are formed from the disintegration of larger cells called megakaryocytes, like that shown in [\[link\]](#)a. For each megakaryocyte, 2000–3000 platelets are formed with 150,000 to 400,000 platelets present in each cubic millimeter of blood. Each platelet is disc shaped and 2–4  $\mu\text{m}$  in diameter. They contain many small vesicles but do not contain a nucleus.



(a)



(b)

(a) Platelets are formed from large cells called megakaryocytes. The megakaryocyte breaks up into thousands of fragments that become platelets. (b) Platelets are required for clotting of the blood. The platelets collect at a wound site in conjunction with other clotting factors, such as fibrinogen, to form a fibrin clot that prevents blood loss and allows the wound to heal.

## Plasma and Serum

The liquid component of blood is called plasma, and it is separated by spinning or centrifuging the blood at high rotations (3000 rpm or higher). The blood cells and platelets are separated by centrifugal forces to the bottom of a specimen tube. The upper liquid layer, the plasma, consists of 90 percent water along with various substances required for maintaining the body's pH, osmotic load, and for protecting the body. The plasma also contains the coagulation factors and antibodies.

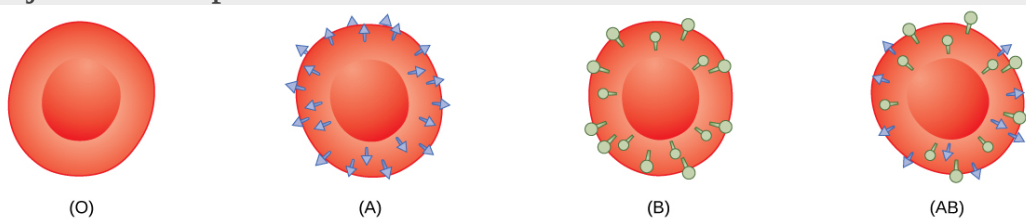
The plasma component of blood without the coagulation factors is called the **serum**. Serum is similar to interstitial fluid in which the correct composition of key ions acting as electrolytes is essential for normal functioning of muscles and nerves. Other components in the serum include proteins that assist with maintaining pH and osmotic balance while giving viscosity to the blood. The serum also contains antibodies, specialized proteins that are important for defense against viruses and bacteria. Lipids, including cholesterol, are also transported in the serum, along with various other substances including nutrients, hormones, metabolic waste, plus external substances, such as, drugs, viruses, and bacteria.

Human serum albumin is the most abundant protein in human blood plasma and is synthesized in the liver. Albumin, which constitutes about half of the blood serum protein, transports hormones and fatty acids, buffers pH, and maintains osmotic pressures. Immunoglobulin is a protein antibody produced in the mucosal lining and plays an important role in antibody mediated immunity.

**Note:****Evolution Connection****Blood Types Related to Proteins on the Surface of the Red Blood Cells**

Red blood cells are coated in antigens made of glycolipids and glycoproteins. The composition of these molecules is determined by genetics, which have evolved over time. In humans, the different surface antigens are grouped into 24 different blood groups with more than 100 different antigens on each red blood cell. The two most well known blood groups are the ABO, shown in [\[link\]](#), and Rh systems. The surface antigens in the ABO blood group are glycolipids, called antigen A and antigen B. People with blood type A have antigen A, those with blood type B have antigen B, those with blood type AB have both antigens, and people with blood type O have neither antigen. Antibodies called agglutinogens are found in the blood plasma and react with the A or B antigens, if the two are mixed. When type A and type B blood are combined, agglutination (clumping) of the blood occurs because of antibodies in the plasma that bind with the opposing antigen; this causes clots that coagulate in the kidney causing kidney failure. Type O blood has neither A or B antigens, and therefore, type O blood can be given to all blood types. Type O negative blood is the universal donor. Type AB positive blood is the universal acceptor because it has both A and B antigen. The ABO blood groups were discovered in 1900 and 1901 by Karl Landsteiner at the University of Vienna.

The Rh blood group was first discovered in Rhesus monkeys. Most people have the Rh antigen (Rh+) and do not have anti-Rh antibodies in their blood. The few people who do not have the Rh antigen and are Rh- can develop anti-Rh antibodies if exposed to Rh+ blood. This can happen after a blood transfusion or after an Rh- woman has an Rh+ baby. The first exposure does not usually cause a reaction; however, at the second exposure, enough antibodies have built up in the blood to produce a reaction that causes agglutination and breakdown of red blood cells. An injection can prevent this reaction.



Human red blood cells may have either type A or B glycoproteins on their surface, both glycoproteins combined (AB), or neither (O). The glycoproteins serve as antigens and can elicit an immune response in a person who receives a transfusion containing unfamiliar antigens. Type O blood, which has no A or B antigens, does not elicit an immune response when injected into a person of any blood type. Thus, O is considered the universal donor. Persons with type AB blood can accept blood from any blood type, and type AB is considered the universal acceptor.

**Note:**

Link to Learning



Play a blood typing game on the [Nobel Prize website](#) to solidify your understanding of blood types.

## Section Summary

Specific components of the blood include red blood cells, white blood cells, platelets, and the plasma, which contains coagulation factors and serum. Blood is important for regulation of the body's pH, temperature, osmotic pressure, the circulation of nutrients and removal of waste, the distribution of hormones from endocrine glands, and the elimination of excess heat; it also contains components for blood clotting. Red blood cells are specialized

cells that contain hemoglobin and circulate through the body delivering oxygen to cells. White blood cells are involved in the immune response to identify and target invading bacteria, viruses, and other foreign organisms; they also recycle waste components, such as old red blood cells. Platelets and blood clotting factors cause the change of the soluble protein fibrinogen to the insoluble protein fibrin at a wound site forming a plug. Plasma consists of 90 percent water along with various substances, such as coagulation factors and antibodies. The serum is the plasma component of the blood without the coagulation factors.

## Review Questions

### Exercise:

**Problem:** White blood cells:

- a. can be classified as granulocytes or agranulocytes
- b. defend the body against bacteria and viruses
- c. are also called leucocytes
- d. All of the above

---

**Solution:**

D

### Exercise:

**Problem:** Platelet plug formation occurs at which point?

- a. when large megakaryocytes break up into thousands of smaller fragments
- b. when platelets are dispersed through the blood stream
- c. when platelets are attracted to a site of blood vessel damage
- d. none of the above



---

**Solution:**

C

**Exercise:**

**Problem:**

In humans, the plasma comprises what percentage of the blood?

- a. 45 percent
- b. 55 percent
- c. 25 percent
- d. 90 percent

---

**Solution:**

B

**Exercise:**

**Problem:**

The red blood cells of birds differ from mammalian red blood cells because:

- a. they are white and have nuclei
- b. they do not have nuclei
- c. they have nuclei
- d. they fight disease

---

**Solution:**

C

**Free Response**

**Exercise:**

**Problem:** Describe the cause of different blood type groups.

---

**Solution:**

Red blood cells are coated with proteins called antigens made of glycolipids and glycoproteins. When type A and type B blood are mixed, the blood agglutinates because of antibodies in the plasma that bind with the opposing antigen. Type O blood has no antigens. The Rh blood group has either the Rh antigen (Rh+) or no Rh antigen (Rh-).

**Exercise:**

**Problem:** List some of the functions of blood in the body.

---

**Solution:**

Blood is important for regulation of the body's pH, temperature, and osmotic pressure, the circulation of nutrients and removal of wastes, the distribution of hormones from endocrine glands, the elimination of excess heat; it also contains components for the clotting of blood to prevent blood loss. Blood also transports clotting factors and disease-fighting agents.

**Exercise:**

**Problem:** How does the lymphatic system work with blood flow?

---

**Solution:**

Lymph capillaries take fluid from the blood to the lymph nodes. The lymph nodes filter the lymph by percolation through connective tissue filled with white blood cells. The white blood cells remove infectious agents, such as bacteria and viruses, to clean the lymph before it returns to the bloodstream.

## Glossary

plasma

liquid component of blood that is left after the cells are removed

platelet

(also, thrombocyte) small cellular fragment that collects at wounds, cross-reacts with clotting factors, and forms a plug to prevent blood loss

red blood cell

small (7–8  $\mu\text{m}$ ) biconcave cell without mitochondria (and in mammals without nuclei) that is packed with hemoglobin, giving the cell its red color; transports oxygen through the body

serum

plasma without the coagulation factors

white blood cell

large (30  $\mu\text{m}$ ) cell with nuclei of which there are many types with different roles including the protection of the body from viruses and bacteria, and cleaning up dead cells and other waste

## Mammalian Heart and Blood Vessels

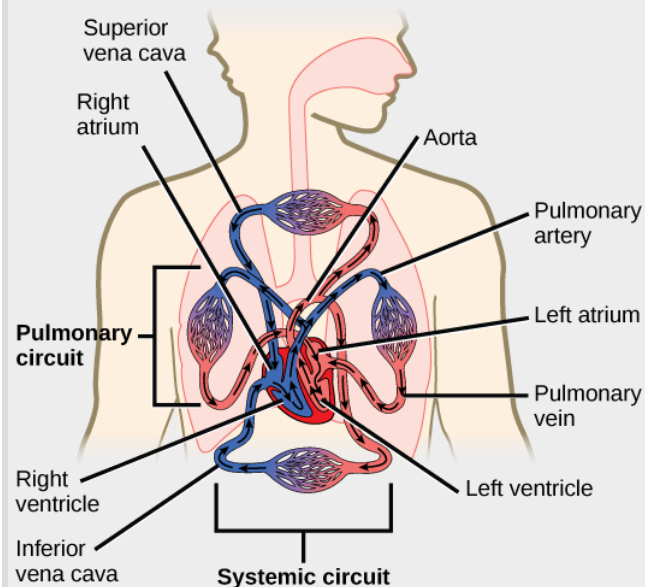
By the end of this section, you will be able to:

- Describe the structure of the heart and explain how cardiac muscle is different from other muscles
- Describe the cardiac cycle
- Explain the structure of arteries, veins, and capillaries, and how blood flows through the body

The heart is a complex muscle that pumps blood through the three divisions of the circulatory system: the coronary (vessels that serve the heart), pulmonary (heart and lungs), and systemic (systems of the body), as shown in [\[link\]](#). Coronary circulation intrinsic to the heart takes blood directly from the main artery (aorta) coming from the heart. For pulmonary and systemic circulation, the heart has to pump blood to the lungs or the rest of the body, respectively. In vertebrates, the lungs are relatively close to the heart in the thoracic cavity. The shorter distance to pump means that the muscle wall on the right side of the heart is not as thick as the left side which must have enough pressure to pump blood all the way to your big toe.

### Note:

#### Art Connection



The mammalian circulatory system is divided into three circuits: the systemic circuit, the pulmonary circuit, and the coronary circuit. Blood is pumped from veins of the systemic circuit into the right atrium of the heart, then into the right ventricle. Blood then enters the pulmonary circuit, and is oxygenated by the lungs. From the pulmonary circuit, blood re-enters the heart through the left atrium. From the left ventricle, blood re-enters the systemic circuit through the aorta and is distributed to the rest of the body.

The coronary circuit, which provides blood to the heart, is not shown.

Which of the following statements about the circulatory system is false?

- a. Blood in the pulmonary vein is deoxygenated.
- b. Blood in the inferior vena cava is deoxygenated.
- c. Blood in the pulmonary artery is deoxygenated.
- d. Blood in the aorta is oxygenated.

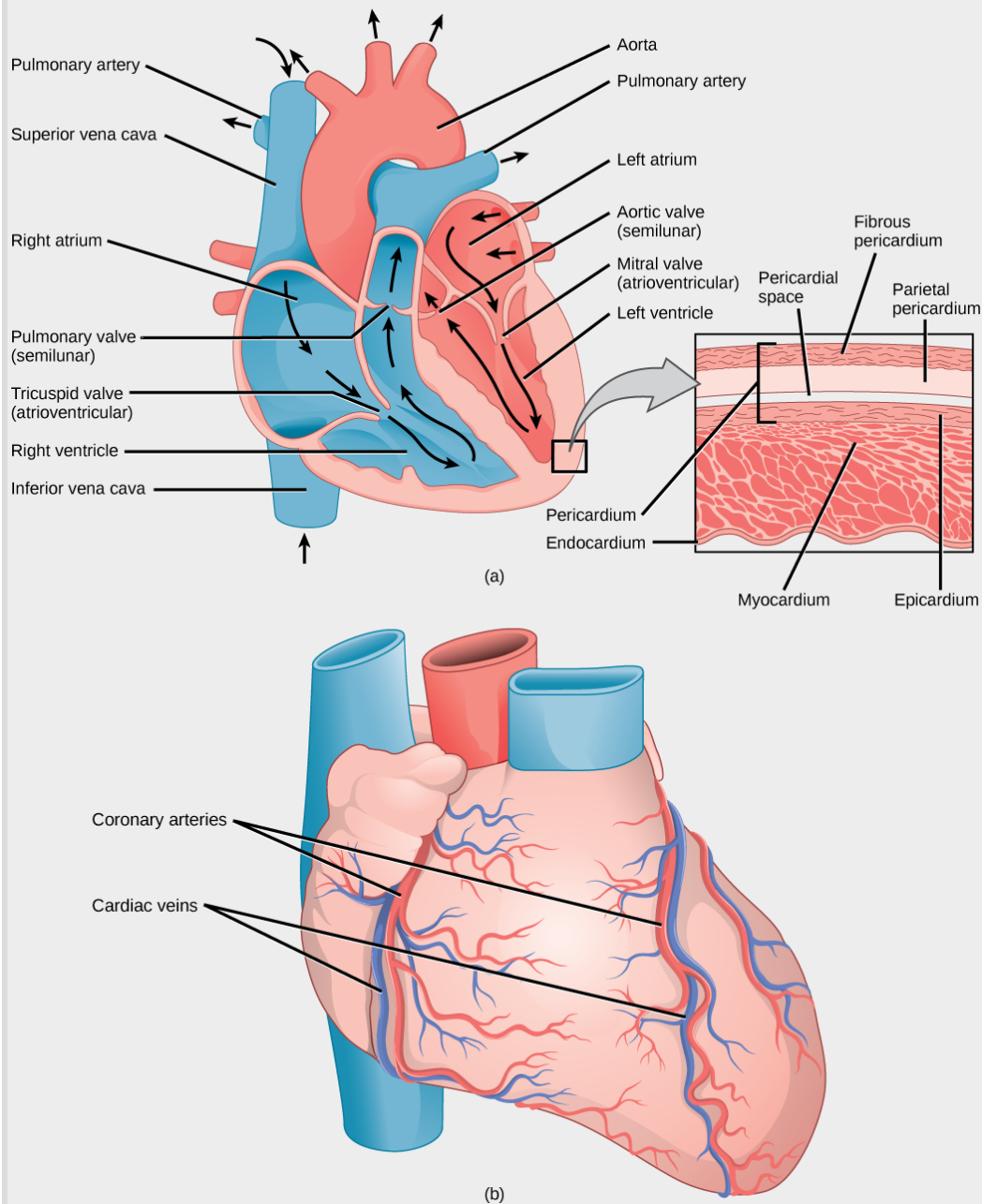
## Structure of the Heart

The heart muscle is asymmetrical as a result of the distance blood must travel in the pulmonary and systemic circuits. Since the right side of the heart sends blood to the pulmonary circuit it is smaller than the left side which must send blood out to the whole body in the systemic circuit, as

shown in [\[link\]](#). In humans, the heart is about the size of a clenched fist; it is divided into four chambers: two atria and two ventricles. There is one atrium and one ventricle on the right side and one atrium and one ventricle on the left side. The atria are the chambers that receive blood, and the ventricles are the chambers that pump blood. The right atrium receives deoxygenated blood from the **superior vena cava**, which drains blood from the jugular vein that comes from the brain and from the veins that come from the arms, as well as from the **inferior vena cava** which drains blood from the veins that come from the lower organs and the legs. In addition, the right atrium receives blood from the coronary sinus which drains deoxygenated blood from the heart itself. This deoxygenated blood then passes to the right ventricle through the **atrioventricular valve** or the **tricuspid valve**, a flap of connective tissue that opens in only one direction to prevent the backflow of blood. The valve separating the chambers on the left side of the heart valve is called the bicuspid or mitral valve. After it is filled, the right ventricle pumps the blood through the pulmonary arteries, by-passing the **semilunar valve** (or pulmonic valve) to the lungs for re-oxygenation. After blood passes through the pulmonary arteries, the right semilunar valves close preventing the blood from flowing backwards into the right ventricle. The left atrium then receives the oxygen-rich blood from the lungs via the pulmonary veins. This blood passes through the **bicuspid valve** or mitral valve (the atrioventricular valve on the left side of the heart) to the left ventricle where the blood is pumped out through **aorta**, the major artery of the body, taking oxygenated blood to the organs and muscles of the body. Once blood is pumped out of the left ventricle and into the aorta, the aortic semilunar valve (or aortic valve) closes preventing blood from flowing backward into the left ventricle. This pattern of pumping is referred to as double circulation and is found in all mammals.

**Note:**

Art Connection



(a) The heart is primarily made of a thick muscle layer, called the myocardium, surrounded by membranes. One-way valves separate the four chambers. (b) Blood vessels of the coronary system, including the coronary arteries and veins, keep the heart musculature oxygenated.

Which of the following statements about the heart is false?

- a. The mitral valve separates the left ventricle from the left atrium.
- b. Blood travels through the bicuspid valve to the left atrium.
- c. Both the aortic and the pulmonary valves are semilunar valves.
- d. The mitral valve is an atrioventricular valve.

The heart is composed of three layers; the epicardium, the myocardium, and the endocardium, illustrated in [\[link\]](#). The inner wall of the heart has a lining called the **endocardium**. The **myocardium** consists of the heart muscle cells that make up the middle layer and the bulk of the heart wall. The outer layer of cells is called the **epicardium**, of which the second layer is a membranous layered structure called the **pericardium** that surrounds and protects the heart; it allows enough room for vigorous pumping but also keeps the heart in place to reduce friction between the heart and other structures.

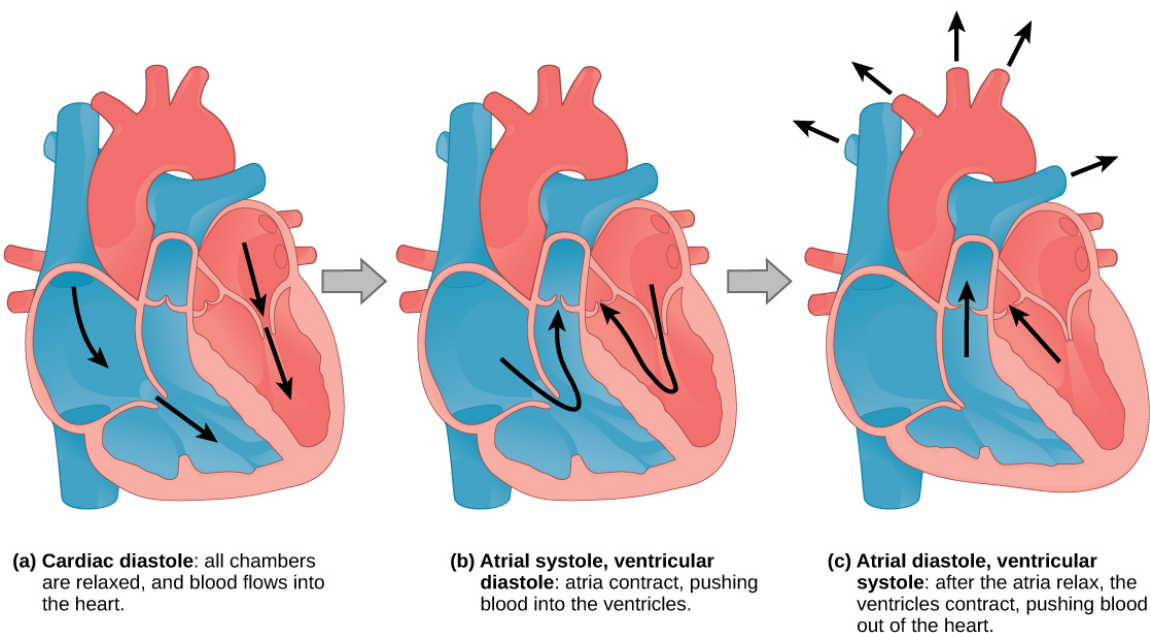
The heart has its own blood vessels that supply the heart muscle with blood. The **coronary arteries** branch from the aorta and surround the outer surface of the heart like a crown. They diverge into capillaries where the heart muscle is supplied with oxygen before converging again into the **coronary veins** to take the deoxygenated blood back to the right atrium where the blood will be re-oxygenated through the pulmonary circuit. The heart muscle will die without a steady supply of blood. **Atherosclerosis** is the blockage of an artery by the buildup of fatty plaques. Because of the size (narrow) of the coronary arteries and their function in serving the heart itself, atherosclerosis can be deadly in these arteries. The slowdown of blood flow and subsequent oxygen deprivation that results from atherosclerosis causes severe pain, known as **angina**, and complete blockage of the arteries will cause **myocardial infarction**: the death of cardiac muscle tissue, commonly known as a heart attack.

## The Cardiac Cycle

The main purpose of the heart is to pump blood through the body; it does so in a repeating sequence called the cardiac cycle. The **cardiac cycle** is the

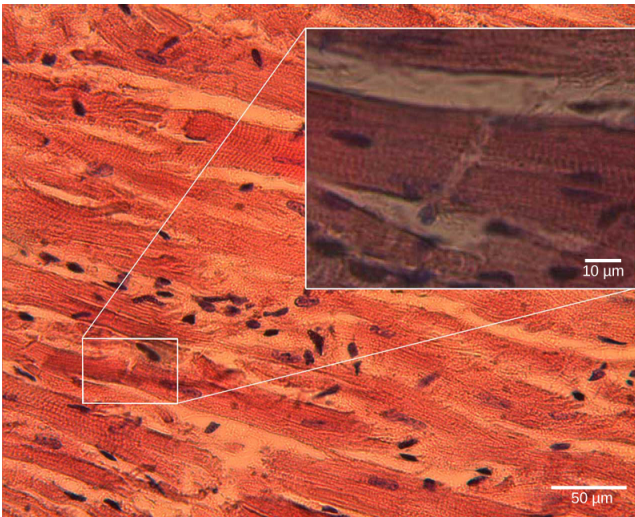


coordination of the filling and emptying of the heart of blood by electrical signals that cause the heart muscles to contract and relax. The human heart beats over 100,000 times per day. In each cardiac cycle, the heart contracts (**systole**), pushing out the blood and pumping it through the body; this is followed by a relaxation phase (**diastole**), where the heart fills with blood, as illustrated in [\[link\]](#). The atria contract at the same time, forcing blood through the atrioventricular valves into the ventricles. Closing of the atrioventricular valves produces a monosyllabic “lup” sound. Following a brief delay, the ventricles contract at the same time forcing blood through the semilunar valves into the aorta and the artery transporting blood to the lungs (via the pulmonary artery). Closing of the semilunar valves produces a monosyllabic “dup” sound.



During (a) cardiac diastole, the heart muscle is relaxed and blood flows into the heart. During (b) atrial systole, the atria contract, pushing blood into the ventricles. During (c) atrial diastole, the ventricles contract, forcing blood out of the heart.

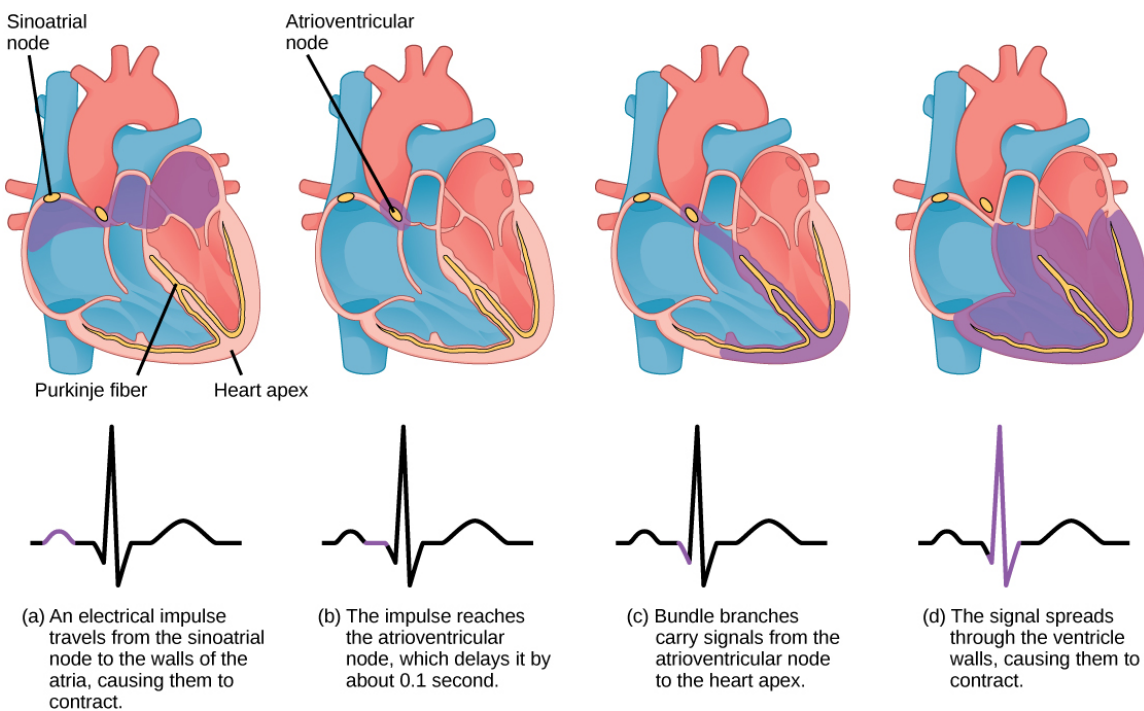
The pumping of the heart is a function of the cardiac muscle cells, or cardiomyocytes, that make up the heart muscle. **Cardiomyocytes**, shown in [\[link\]](#), are distinctive muscle cells that are striated like skeletal muscle but pump rhythmically and involuntarily like smooth muscle; they are connected by intercalated disks exclusive to cardiac muscle. They are self-stimulated for a period of time and isolated cardiomyocytes will beat if given the correct balance of nutrients and electrolytes.



Cardiomyocytes are striated muscle cells found in cardiac tissue. (credit: modification of work by Dr. S. Girod, Anton Becker; scale-bar data from Matt Russell)

The autonomous beating of cardiac muscle cells is regulated by the heart's internal pacemaker that uses electrical signals to time the beating of the heart. The electrical signals and mechanical actions, illustrated in [\[link\]](#), are intimately intertwined. The internal pacemaker starts at the **sinoatrial (SA) node**, which is located near the wall of the right atrium. Electrical charges spontaneously pulse from the SA node causing the two atria to contract in

unison. The pulse reaches a second node, called the atrioventricular (AV) node, between the right atrium and right ventricle where it pauses for approximately 0.1 second before spreading to the walls of the ventricles. From the AV node, the electrical impulse enters the bundle of His, then to the left and right bundle branches extending through the interventricular septum. Finally, the Purkinje fibers conduct the impulse from the apex of the heart up the ventricular myocardium, and then the ventricles contract. This pause allows the atria to empty completely into the ventricles before the ventricles pump out the blood. The electrical impulses in the heart produce electrical currents that flow through the body and can be measured on the skin using electrodes. This information can be observed as an **electrocardiogram (ECG)**—a recording of the electrical impulses of the cardiac muscle.



The beating of the heart is regulated by an electrical impulse that causes the characteristic reading of an ECG. The signal is initiated at the sinoatrial valve. The signal then (a) spreads to the atria, causing them to contract. The signal is (b) delayed at the atrioventricular node before it is passed on to the (c) heart apex.

The delay allows the atria to relax before the (d) ventricles contract. The final part of the ECG cycle prepares the heart for the next beat.

**Note:**

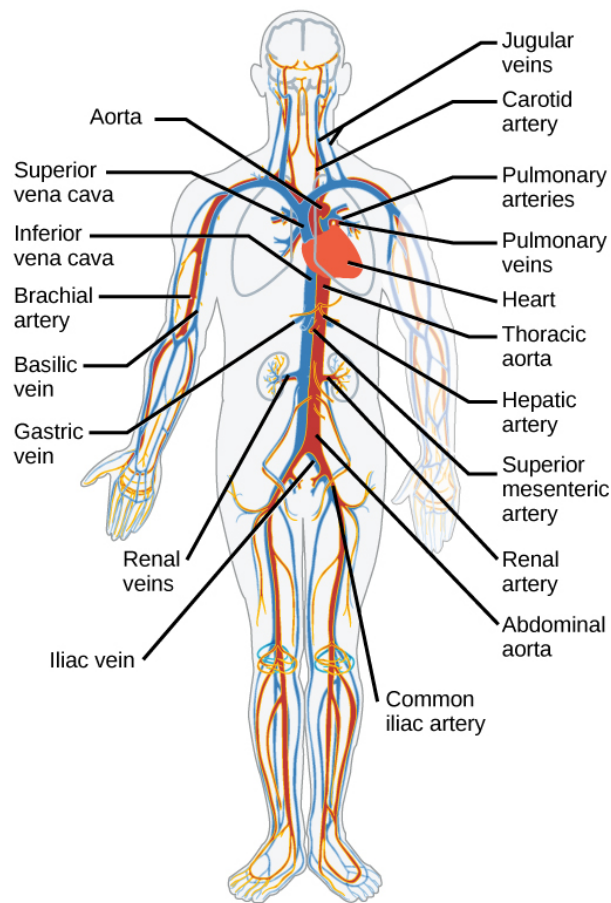
Link to Learning



Visit [this site](#) and select the dropdown “Your Heart’s Electrical System” to see the heart’s “pacemaker” in action.

## Arteries, Veins, and Capillaries

The blood from the heart is carried through the body by a complex network of blood vessels ([\[link\]](#)). **Arteries** take blood away from the heart. The main artery is the aorta that branches into major arteries that take blood to different limbs and organs. These major arteries include the carotid artery that takes blood to the brain, the brachial arteries that take blood to the arms, and the thoracic artery that takes blood to the thorax and then into the hepatic, renal, and gastric arteries for the liver, kidney, and stomach, respectively. The iliac artery takes blood to the lower limbs. The major arteries diverge into minor arteries, and then smaller vessels called **arterioles**, to reach more deeply into the muscles and organs of the body.



The major human arteries and veins are shown. (credit: modification of work by Mariana Ruiz Villareal)

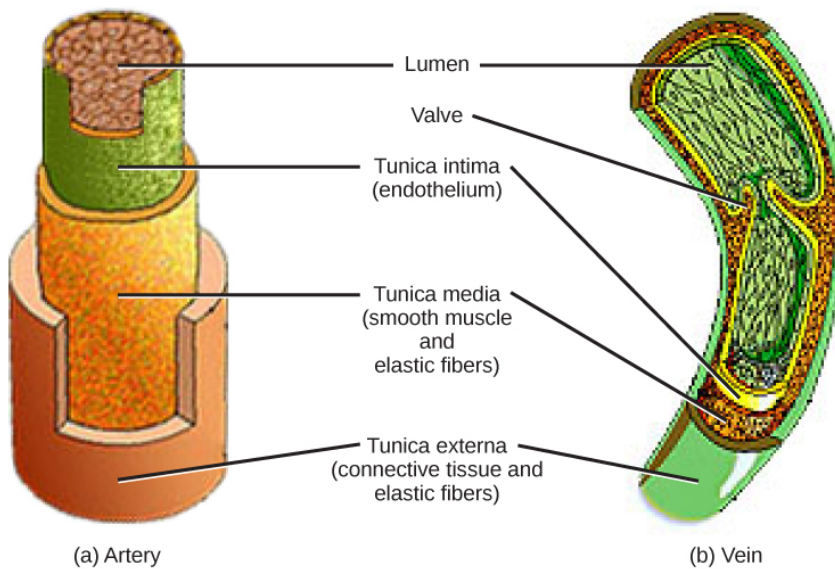
Arterioles diverge into capillary beds. **Capillary beds** contain a large number (10 to 100) of **capillaries** that branch among the cells and tissues of the body. Capillaries are narrow-diameter tubes that can fit red blood cells through in single file and are the sites for the exchange of nutrients, waste, and oxygen with tissues at the cellular level. Fluid also crosses into the interstitial space from the capillaries. The capillaries converge again into **venules** that connect to minor veins that finally connect to major veins that take blood high in carbon dioxide back to the heart. **Veins** are blood vessels that bring blood back to the heart. The major veins drain blood from the

same organs and limbs that the major arteries supply. Fluid is also brought back to the heart via the lymphatic system.

The structure of the different types of blood vessels reflects their function or layers. There are three distinct layers, or tunics, that form the walls of blood vessels ([\[link\]](#)). The first tunic is a smooth, inner lining of endothelial cells that are in contact with the red blood cells. The endothelial tunic is continuous with the endocardium of the heart. In capillaries, this single layer of cells is the location of diffusion of oxygen and carbon dioxide between the endothelial cells and red blood cells, as well as the exchange site via endocytosis and exocytosis. The movement of materials at the site of capillaries is regulated by **vasoconstriction**, narrowing of the blood vessels, and **vasodilation**, widening of the blood vessels; this is important in the overall regulation of blood pressure.

Veins and arteries both have two further tunics that surround the endothelium: the middle tunic is composed of smooth muscle and the outermost layer is connective tissue (collagen and elastic fibers). The elastic connective tissue stretches and supports the blood vessels, and the smooth muscle layer helps regulate blood flow by altering vascular resistance through vasoconstriction and vasodilation. The arteries have thicker smooth muscle and connective tissue than the veins to accommodate the higher pressure and speed of freshly pumped blood. The veins are thinner walled as the pressure and rate of flow are much lower. In addition, veins are structurally different than arteries in that veins have valves to prevent the backflow of blood. Because veins have to work against gravity to get blood back to the heart, contraction of skeletal muscle assists with the flow of blood back to the heart.





Arteries and veins consist of three layers: an outer tunica externa, a middle tunica media, and an inner tunica intima. Capillaries consist of a single layer of epithelial cells, the tunica intima. (credit: modification of work by NCI, NIH)

## Section Summary

The heart muscle pumps blood through three divisions of the circulatory system: coronary, pulmonary, and systemic. There is one atrium and one ventricle on the right side and one atrium and one ventricle on the left side. The pumping of the heart is a function of cardiomyocytes, distinctive muscle cells that are striated like skeletal muscle but pump rhythmically and involuntarily like smooth muscle. The internal pacemaker starts at the sinoatrial node, which is located near the wall of the right atrium. Electrical charges pulse from the SA node causing the two atria to contract in unison; then the pulse reaches the atrioventricular node between the right atrium and right ventricle. A pause in the electric signal allows the atria to empty completely into the ventricles before the ventricles pump out the blood. The blood from the heart is carried through the body by a complex network of

blood vessels; arteries take blood away from the heart, and veins bring blood back to the heart.

## Art Connections

### Exercise:

#### Problem:

[\[link\]](#) Which of the following statements about the circulatory system is false?

- a. Blood in the pulmonary vein is deoxygenated.
- b. Blood in the inferior vena cava is deoxygenated.
- c. Blood in the pulmonary artery is deoxygenated.
- d. Blood in the aorta is oxygenated.

---

#### Solution:

[\[link\]](#) C

### Exercise:

#### Problem:

[\[link\]](#) Which of the following statements about the heart is false?

- a. The mitral valve separates the left ventricle from the left atrium.
- b. Blood travels through the bicuspid valve to the left atrium.
- c. Both the aortic and the pulmonary valves are semilunar valves.
- d. The mitral valve is an atrioventricular valve.

---

#### Solution:

[\[link\]](#) B

## Review Questions



**Exercise:**

**Problem:** The heart's internal pacemaker beats by:

- a. an internal implant that sends an electrical impulse through the heart
  - b. the excitation of cardiac muscle cells at the sinoatrial node followed by the atrioventricular node
  - c. the excitation of cardiac muscle cells at the atrioventricular node followed by the sinoatrial node
  - d. the action of the sinus
- 

**Solution:**

B

**Exercise:**

**Problem:**

During the systolic phase of the cardiac cycle, the heart is \_\_\_\_\_.

- a. contracting
  - b. relaxing
  - c. contracting and relaxing
  - d. filling with blood
- 

**Solution:**

A

**Exercise:**

**Problem:** Cardiomyocytes are similar to skeletal muscle because:

- a. they beat involuntarily
- b. they are used for weight lifting

- c. they pulse rhythmically
- d. they are striated

---

**Solution:**

D

**Exercise:**

**Problem:**How do arteries differ from veins?

- a. Arteries have thicker smooth muscle layers to accommodate the changes in pressure from the heart.
- b. Arteries carry blood.
- c. Arteries have thinner smooth muscle layers and valves and move blood by the action of skeletal muscle.
- d. Arteries are thin walled and are used for gas exchange.

---

**Solution:**

A

**Free Response**

**Exercise:**

**Problem:**Describe the cardiac cycle.

---

**Solution:**

The heart receives an electrical signal from the sinoatrial node triggering the cardiac muscle cells in the atria to contract. The signal pauses at the atrioventricular node before spreading to the walls of the ventricles so the blood is pumped through the body. This is the systolic phase. The heart then relaxes in the diastole and fills again with blood.

## Exercise:

**Problem:**What happens in capillaries?

---

### Solution:

The capillaries basically exchange materials with their surroundings. Their walls are very thin and are made of one or two layers of cells, where gases, nutrients, and waste are diffused. They are distributed as beds, complex networks that link arteries as well as veins.

## Glossary

angina

pain caused by partial blockage of the coronary arteries by the buildup of plaque and lack of oxygen to the heart muscle

aorta

major artery of the body that takes blood away from the heart

arteriole

small vessel that connects an artery to a capillary bed

artery

blood vessel that takes blood away from the heart

atherosclerosis

buildup of fatty plaques in the coronary arteries in the heart

atrioventricular valve

one-way membranous flap of connective tissue between the atrium and the ventricle in the right side of the heart; also known as tricuspid valve

bicuspid valve

(also, mitral valve; left atrioventricular valve) one-way membranous flap between the atrium and the ventricle in the left side of the heart

capillary

smallest blood vessel that allows the passage of individual blood cells and the site of diffusion of oxygen and nutrient exchange

capillary bed

large number of capillaries that converge to take blood to a particular organ or tissue

cardiac cycle

filling and emptying the heart of blood by electrical signals that cause the heart muscles to contract and relax

cardiomyocyte

specialized heart muscle cell that is striated but contracts involuntarily like smooth muscle

coronary artery

vessel that supplies the heart tissue with blood

coronary vein

vessel that takes blood away from the heart tissue back to the chambers in the heart

diastole

relaxation phase of the cardiac cycle when the heart is relaxed and the ventricles are filling with blood

electrocardiogram (ECG)

recording of the electrical impulses of the cardiac muscle

endocardium

innermost layer of tissue in the heart

epicardium

outermost tissue layer of the heart

inferior vena cava

drains blood from the veins that come from the lower organs and the legs

myocardial infarction

(also, heart attack) complete blockage of the coronary arteries and death of the cardiac muscle tissue

myocardium

heart muscle cells that make up the middle layer and the bulk of the heart wall

pericardium

membrane layer protecting the heart; also part of the epicardium

semilunar valve

membranous flap of connective tissue between the aorta and a ventricle of the heart (the aortic or pulmonary semilunar valves)

sinoatrial (SA) node

the heart's internal pacemaker; located near the wall of the right atrium

superior vena cava

drains blood from the jugular vein that comes from the brain and from the veins that come from the arms

systole

contraction phase of cardiac cycle when the ventricles are pumping blood into the arteries

tricuspid valve

one-way membranous flap of connective tissue between the atrium and the ventricle in the right side of the heart; also known as atrioventricular valve

vasoconstriction

narrowing of a blood vessel

vasodilation

widening of a blood vessel

vein

blood vessel that brings blood back to the heart

vena cava

major vein of the body returning blood from the upper and lower parts of the body; see the superior vena cava and inferior vena cava

venule

blood vessel that connects a capillary bed to a vein

## Blood Flow and Blood Pressure Regulation

By the end of this section, you will be able to:

- Describe the system of blood flow through the body
- Describe how blood pressure is regulated

**Blood pressure (BP)** is the pressure exerted by blood on the walls of a blood vessel that helps to push blood through the body. Systolic blood pressure measures the amount of pressure that blood exerts on vessels while the heart is beating. The optimal systolic blood pressure is 120 mmHg. Diastolic blood pressure measures the pressure in the vessels between heartbeats. The optimal diastolic blood pressure is 80 mmHg. Many factors can affect blood pressure, such as hormones, stress, exercise, eating, sitting, and standing. Blood flow through the body is regulated by the size of blood vessels, by the action of smooth muscle, by one-way valves, and by the fluid pressure of the blood itself.

## How Blood Flows Through the Body

Blood is pushed through the body by the action of the pumping heart. With each rhythmic pump, blood is pushed under high pressure and velocity away from the heart, initially along the main artery, the aorta. In the aorta, the blood travels at 30 cm/sec. As blood moves into the arteries, arterioles, and ultimately to the capillary beds, the rate of movement slows dramatically to about 0.026 cm/sec, one-thousand times slower than the rate of movement in the aorta. While the diameter of each individual arteriole and capillary is far narrower than the diameter of the aorta, and according to the law of continuity, fluid should travel faster through a narrower diameter tube, the rate is actually slower due to the overall diameter of all the combined capillaries being far greater than the diameter of the individual aorta.

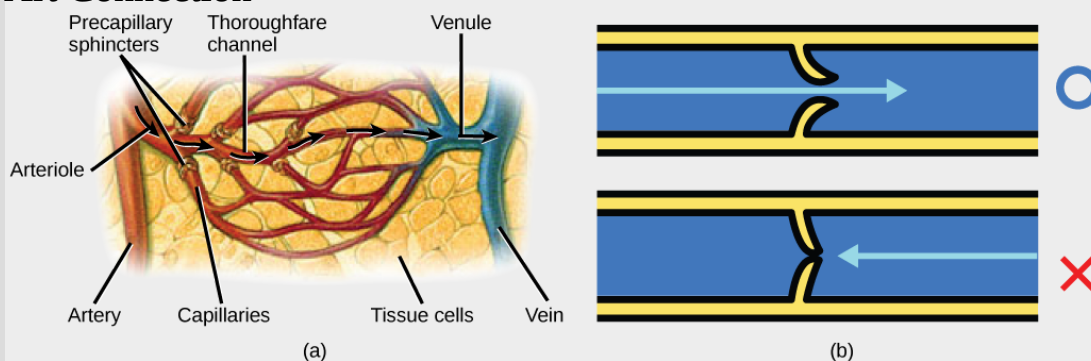
The slow rate of travel through the capillary beds, which reach almost every cell in the body, assists with gas and nutrient exchange and also promotes the diffusion of fluid into the interstitial space. After the blood has passed through the capillary beds to the venules, veins, and finally to the main venae cavae, the rate of flow increases again but is still much slower than

the initial rate in the aorta. Blood primarily moves in the veins by the rhythmic movement of smooth muscle in the vessel wall and by the action of the skeletal muscle as the body moves. Because most veins must move blood against the pull of gravity, blood is prevented from flowing backward in the veins by one-way valves. Because skeletal muscle contraction aids in venous blood flow, it is important to get up and move frequently after long periods of sitting so that blood will not pool in the extremities.

Blood flow through the capillary beds is regulated depending on the body's needs and is directed by nerve and hormone signals. For example, after a large meal, most of the blood is diverted to the stomach by vasodilation of vessels of the digestive system and vasoconstriction of other vessels. During exercise, blood is diverted to the skeletal muscles through vasodilation while blood to the digestive system would be lessened through vasoconstriction. The blood entering some capillary beds is controlled by small muscles, called precapillary sphincters, illustrated in [\[link\]](#). If the sphincters are open, the blood will flow into the associated branches of the capillary blood. If all of the sphincters are closed, then the blood will flow directly from the arteriole to the venule through the thoroughfare channel (see [\[link\]](#)). These muscles allow the body to precisely control when capillary beds receive blood flow. At any given moment only about 5-10% of our capillary beds actually have blood flowing through them.

### Note:

#### Art Connection



(a) Precapillary sphincters are rings of smooth muscle that regulate the flow of blood through capillaries; they help control



the location of blood flow to where it is needed. (b) Valves in the veins prevent blood from moving backward. (credit a: modification of work by NCI)

Varicose veins are veins that become enlarged because the valves no longer close properly, allowing blood to flow backward. Varicose veins are often most prominent on the legs. Why do you think this is the case?

**Note:**

Link to Learning



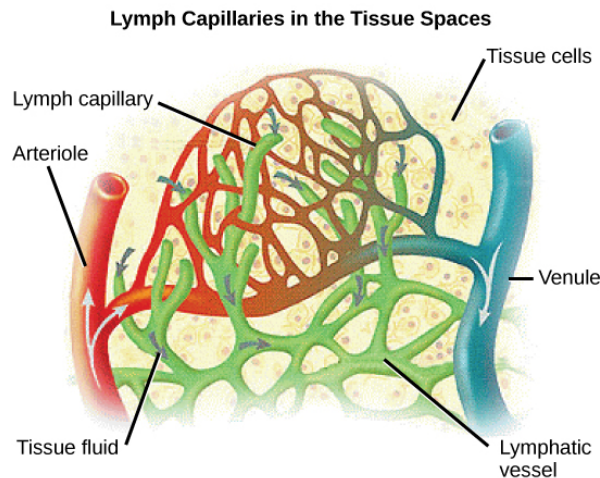
See the circulatory system's blood flow.

<https://www.openstaxcollege.org/l/circulation>

Proteins and other large solutes cannot leave the capillaries. The loss of the watery plasma creates a hyperosmotic solution within the capillaries, especially near the venules. This causes about 85% of the plasma that leaves the capillaries to eventually diffuse back into the capillaries near the venules. The remaining 15% of blood plasma drains out from the interstitial fluid into nearby lymphatic vessels ([link]). The fluid in the lymph is similar in composition to the interstitial fluid. The lymph fluid passes through lymph nodes before it returns to the heart via the vena cava.

**Lymph nodes** are specialized organs that filter the lymph by percolation through a maze of connective tissue filled with white blood cells. The white blood cells remove infectious agents, such as bacteria and viruses, to clean the lymph before it returns to the bloodstream. After it is cleaned, the lymph

returns to the heart by the action of smooth muscle pumping, skeletal muscle action, and one-way valves joining the returning blood near the junction of the venae cavae entering the right atrium of the heart.



Fluid from the capillaries moves into the interstitial space and lymph capillaries by diffusion down a pressure gradient and also by osmosis.

Out of 7,200 liters of fluid pumped by the average heart in a day, over 1,500 liters is filtered. (credit: modification of work by NCI, NIH)

**Note:**

**Evolution Connection**

**Vertebrate Diversity in Blood Circulation**

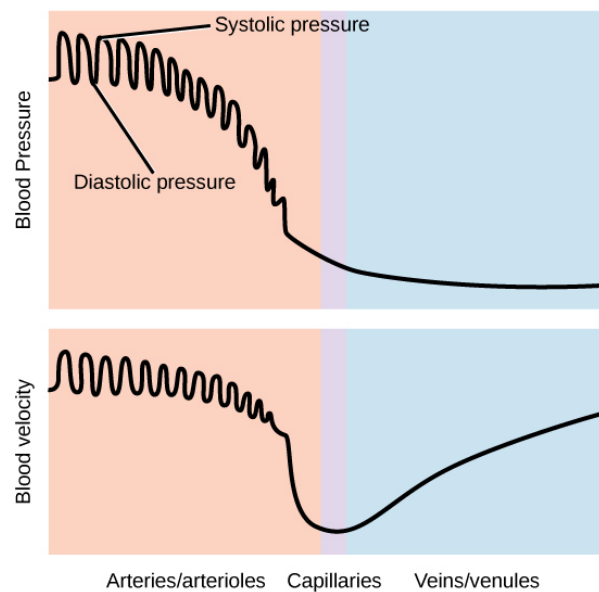
Blood circulation has evolved differently in vertebrates and may show variation in different animals for the required amount of pressure, organ

and vessel location, and organ size. Animals with long necks and those that live in cold environments have distinct blood pressure adaptations. Long-necked animals, such as giraffes, need to pump blood upward from the heart against gravity. The blood pressure required from the pumping of the left ventricle would be equivalent to 250 mm Hg (mm Hg = millimeters of mercury, a unit of pressure) to reach the height of a giraffe's head, which is 2.5 meters higher than the heart. However, if checks and balances were not in place, this blood pressure would damage the giraffe's brain, particularly if it was bending down to drink. These checks and balances include valves and feedback mechanisms that reduce the rate of cardiac output. Long-necked dinosaurs such as the sauropods had to pump blood even higher, up to ten meters above the heart. This would have required a blood pressure of more than 600 mm Hg, which could only have been achieved by an enormous heart. Evidence for such an enormous heart does not exist and mechanisms to reduce the blood pressure required include the slowing of metabolism as these animals grew larger. It is likely that they did not routinely feed on tree tops but grazed on the ground. Living in cold water, whales need to maintain the temperature in their blood. This is achieved by the veins and arteries being close together so that heat exchange can occur. This mechanism is called a countercurrent heat exchanger. The blood vessels and the whole body are also protected by thick layers of blubber to prevent heat loss. In land animals that live in cold environments, thick fur and hibernation are used to retain heat and slow metabolism.

## **Blood Pressure**

The pressure of the blood flow in the body is produced by the hydrostatic pressure of the fluid (blood) against the walls of the blood vessels. Fluid will move from areas of high to low hydrostatic pressures. In the arteries, the hydrostatic pressure near the heart is very high and blood flows to the arterioles where the rate of flow is slowed by the narrow openings of the arterioles. During systole, when new blood is entering the arteries, the artery walls stretch to accommodate the increase of pressure of the extra blood; during diastole, the walls return to normal because of their elastic

properties. The blood pressure of the systole phase and the diastole phase, graphed in [\[link\]](#), gives the two pressure readings for blood pressure. For example, 120/80 indicates a reading of 120 mm Hg during the systole and 80 mm Hg during diastole. Throughout the cardiac cycle, the blood continues to empty into the arterioles at a relatively even rate. This resistance to blood flow is called **peripheral resistance**.



Blood pressure is related to the blood velocity in the arteries and arterioles. In the capillaries and veins, the blood pressure continues to decrease but velocity increases.

## Blood Pressure Regulation

Cardiac output is the volume of blood pumped by the heart in one minute. It is calculated by multiplying the number of heart contractions that occur per minute (heart rate) times the **stroke volume** (the volume of blood pumped

into the aorta per contraction of the left ventricle). Therefore, cardiac output can be increased by increasing heart rate, as when exercising. However, cardiac output can also be increased by increasing stroke volume, such as if the heart contracts with greater strength. Stroke volume can also be increased by speeding blood circulation through the body so that more blood enters the heart between contractions. During heavy exertion, the blood vessels relax and increase in diameter, offsetting the increased heart rate and ensuring adequate oxygenated blood gets to the muscles. Stress triggers a decrease in the diameter of the blood vessels, consequently increasing blood pressure. These changes can also be caused by nerve signals or hormones, and even standing up or lying down can have a great effect on blood pressure.

## **Section Summary**

Blood primarily moves through the body by the rhythmic movement of smooth muscle in the vessel wall and by the action of the skeletal muscle as the body moves. Blood is prevented from flowing backward in the veins by one-way valves. Blood flow through the capillary beds is controlled by precapillary sphincters to increase and decrease flow depending on the body's needs and is directed by nerve and hormone signals. Lymph vessels take fluid that has leaked out of the blood to the lymph nodes where it is cleaned before returning to the heart. During systole, blood enters the arteries, and the artery walls stretch to accommodate the extra blood. During diastole, the artery walls return to normal. The blood pressure of the systole phase and the diastole phase gives the two pressure readings for blood pressure.

## **Art Connections**

### **Exercise:**

**Problem:**

[\[link\]](#) Varicose veins are veins that become enlarged because the valves no longer close properly, allowing blood to flow backward. Varicose veins are often most prominent on the legs. Why do you think this is the case?

---

**Solution:**

[\[link\]](#) Blood in the legs is farthest away from the heart and has to flow up to reach it.

**Review Questions****Exercise:**

**Problem:** High blood pressure would be a result of \_\_\_\_\_.

- a. a high cardiac output and high peripheral resistance
  - b. a high cardiac output and low peripheral resistance
  - c. a low cardiac output and high peripheral resistance
  - d. a low cardiac output and low peripheral resistance
- 

**Solution:**

A

**Free Response****Exercise:**

**Problem:** How does blood pressure change during heavy exercise?

---

**Solution:**

The heart rate increases, which increases the hydrostatic pressure against the artery walls. At the same time, the arterioles dilate in response to the increased exercise, which reduces peripheral resistance.

## **Glossary**

blood pressure (BP)

pressure of blood in the arteries that helps to push blood through the body

cardiac output

the volume of blood pumped by the heart in one minute as a product of heart rate multiplied by stroke volume

lymph node

specialized organ that contains a large number of macrophages that clean the lymph before the fluid is returned to the heart

peripheral resistance

resistance of the artery and blood vessel walls to the pressure placed on them by the force of the heart pumping

precapillary sphincter

small muscle that controls blood circulation in the capillary beds

stroke volume>

- the volume of blood pumped into the aorta per contraction of the left ventricle

## Introduction

class="introduction"

Just as  
humans  
recycle what  
we can and  
dump the  
remains into  
landfills, our  
bodies use  
and recycle  
what they  
can and  
excrete the  
remaining  
waste  
products.  
Our bodies'  
complex  
systems  
have  
developed  
ways to treat  
waste and  
maintain a  
balanced  
internal  
environment  
. (credit:  
modification  
of work by  
Redwin  
Law)





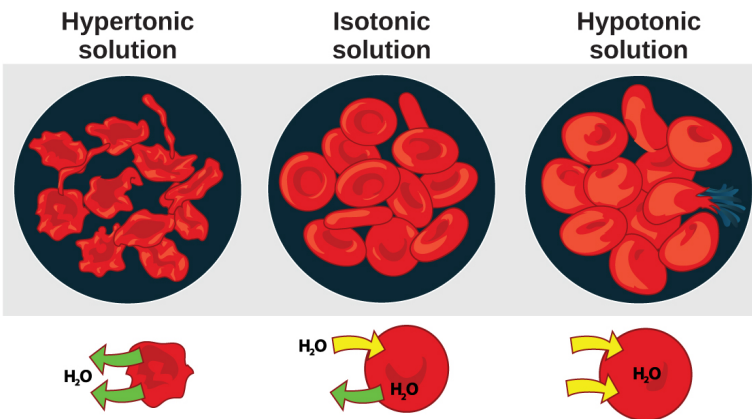
The daily intake recommendation for human water consumption is eight to ten glasses of water. In order to achieve a healthy balance, the human body should excrete the eight to ten glasses of water every day. This occurs via the processes of urination, defecation, sweating and, to a small extent, respiration. The organs and tissues of the human body are soaked in fluids that are maintained at constant temperature, pH, and solute concentration, all crucial elements of homeostasis. The solutes in body fluids are mainly mineral salts and sugars, and osmotic regulation is the process by which the mineral salts and water are kept in balance. Osmotic homeostasis is maintained despite the influence of external factors like temperature, diet, and weather conditions.

## Osmoregulation and Osmotic Balance

By the end of this section, you will be able to:

- Define osmosis and explain its role within molecules
- Explain why osmoregulation and osmotic balance are important body functions
- Describe active transport mechanisms
- Explain osmolarity and the way in which it is measured
- Describe osmoregulators or osmoconformers and how these tools allow animals to adapt to different environments

Osmosis is the diffusion of water across a membrane in response to **osmotic pressure** caused by an imbalance of molecules on either side of the membrane. **Osmoregulation** is the process of maintenance of salt and water balance (**osmotic balance**) across membranes within the body's fluids, which are composed of water, plus electrolytes and non-electrolytes. An **electrolyte** is a solute that dissociates into ions when dissolved in water. A **non-electrolyte**, in contrast, doesn't dissociate into ions during water dissolution. Both electrolytes and non-electrolytes contribute to the osmotic balance. The body's fluids include blood plasma, the cytosol within cells, and interstitial fluid, the fluid that exists in the spaces between cells and tissues of the body. The membranes of the body (such as the pleural, serous, and cell membranes) are **semi-permeable membranes**. Semi-permeable membranes are permeable (or permissive) to certain types of solutes and water. Solutions on two sides of a semi-permeable membrane tend to equalize in solute concentration by movement of solutes and/or water across the membrane. As seen in [\[link\]](#), a cell placed in water tends to swell due to gain of water from the hypotonic or "low salt" environment. A cell placed in a solution with higher salt concentration, on the other hand, tends to make the membrane shrivel up due to loss of water into the hypertonic or "high salt" environment. Isotonic cells have an equal concentration of solutes inside and outside the cell; this equalizes the osmotic pressure on either side of the cell membrane which is a semi-permeable membrane.



Cells placed in a hypertonic environment tend to shrink due to loss of water. In a hypotonic environment, cells tend to swell due to intake of water. The blood maintains an isotonic environment so that cells neither shrink nor swell. (credit: Mariana Ruiz Villareal)

The body does not exist in isolation. There is a constant input of water and electrolytes into the system. While osmoregulation is achieved across membranes within the body, excess electrolytes and wastes are transported to the kidneys and excreted, helping to maintain osmotic balance.

## Need for Osmoregulation

Biological systems constantly interact and exchange water and nutrients with the environment by way of consumption of food and water and through excretion in the form of sweat, urine, and feces. Without a mechanism to regulate osmotic pressure, or when a disease damages this mechanism, there is a tendency to accumulate toxic waste and water, which can have dire consequences.

Mammalian systems have evolved to regulate not only the overall osmotic pressure across membranes, but also specific concentrations of important electrolytes in the three major fluid compartments: blood plasma,

extracellular fluid, and intracellular fluid. Since osmotic pressure is regulated by the movement of water across membranes, the volume of the fluid compartments can also change temporarily. Because blood plasma is one of the fluid components, osmotic pressures have a direct bearing on blood pressure.

## **Transport of Electrolytes across Cell Membranes**

Electrolytes, such as sodium chloride, ionize in water, meaning that they dissociate into their component ions. In water, sodium chloride ( $\text{NaCl}$ ), dissociates into the sodium ion ( $\text{Na}^+$ ) and the chloride ion ( $\text{Cl}^-$ ). The most important ions, whose concentrations are very closely regulated in body fluids, are the cations sodium ( $\text{Na}^+$ ), potassium ( $\text{K}^+$ ), calcium ( $\text{Ca}^{+2}$ ), magnesium ( $\text{Mg}^{+2}$ ), and the anions chloride ( $\text{Cl}^-$ ), carbonate ( $\text{CO}_3^{-2}$ ), bicarbonate ( $\text{HCO}_3^-$ ), and phosphate ( $\text{PO}_3^-$ ). Electrolytes are lost from the body during urination and perspiration. For this reason, athletes are encouraged to replace electrolytes and fluids during periods of increased activity and perspiration.

Osmotic pressure is influenced by the concentration of solutes in a solution. It is directly proportional to the number of solute atoms or molecules and not dependent on the size of the solute molecules. Because electrolytes dissociate into their component ions, they, in essence, add more solute particles into the solution and have a greater effect on osmotic pressure, per mass than compounds that do not dissociate in water, such as glucose.

Water can pass through membranes by passive diffusion. If electrolyte ions could passively diffuse across membranes, it would be impossible to maintain specific concentrations of ions in each fluid compartment therefore they require special mechanisms to cross the semi-permeable membranes in the body. This movement can be accomplished by facilitated diffusion and active transport. Facilitated diffusion requires protein-based channels for moving the solute. Active transport requires energy in the form of ATP conversion, carrier proteins, or pumps in order to move ions against the concentration gradient.

## **Concept of Osmolality and Milliequivalent**

In order to calculate osmotic pressure, it is necessary to understand how solute concentrations are measured. The unit for measuring solutes is the **mole**. One mole is defined as the gram molecular weight of the solute. For example, the molecular weight of sodium chloride is 58.44. Thus, one mole of sodium chloride weighs 58.44 grams. The **molarity** of a solution is the number of moles of solute per liter of solution. The **molality** of a solution is the number of moles of solute per kilogram of solvent. If the solvent is water, one kilogram of water is equal to one liter of water. While molarity and molality are used to express the concentration of solutions, electrolyte concentrations are usually expressed in terms of milliequivalents per liter (mEq/L): the mEq/L is equal to the ion concentration (in millimoles) multiplied by the number of electrical charges on the ion. The unit of milliequivalent takes into consideration the ions present in the solution (since electrolytes form ions in aqueous solutions) and the charge on the ions.

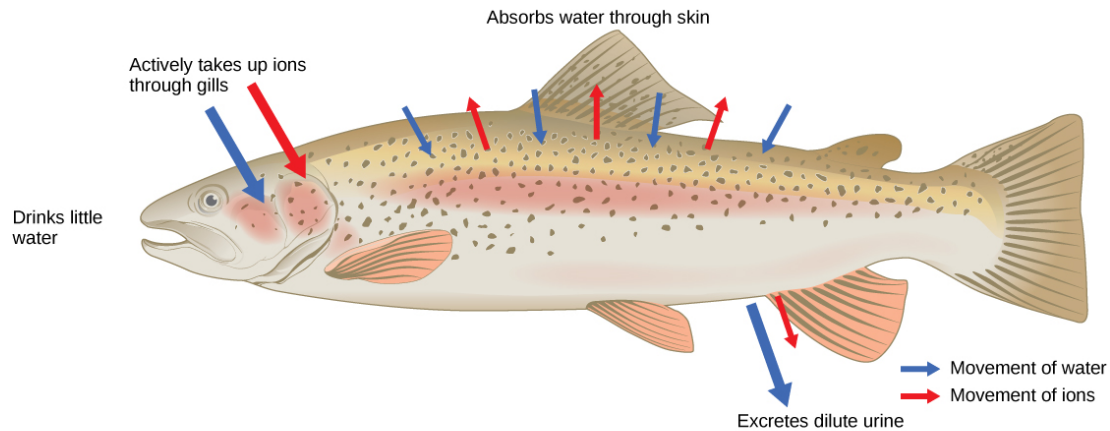
Thus, for ions that have a charge of one, one milliequivalent is equal to one millimole. For ions that have a charge of two (like calcium), one milliequivalent is equal to 0.5 millimoles. Another unit for the expression of electrolyte concentration is the milliosmole (mOsm), which is the number of milliequivalents of solute per kilogram of solvent. Body fluids are usually maintained within the range of 280 to 300 mOsm.

## **Osmoregulators and Osmoconformers**

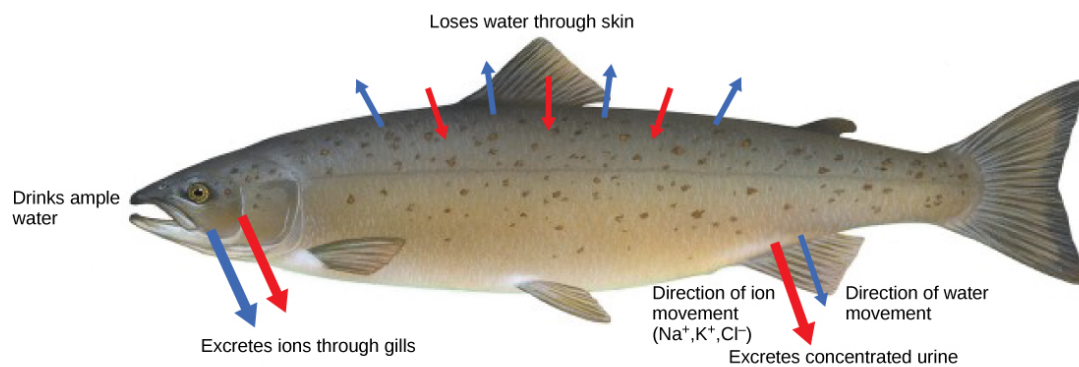
Persons lost at sea without any fresh water to drink are at risk of severe dehydration because the human body cannot adapt to drinking seawater, which is hypertonic in comparison to body fluids. Organisms such as goldfish that can tolerate only a relatively narrow range of salinity are referred to as stenohaline. About 90 percent of all bony fish are restricted to either freshwater or seawater. They are incapable of osmotic regulation in the opposite environment. It is possible, however, for a few fishes like salmon to spend part of their life in fresh water and part in sea water. Organisms like the salmon and molly that can tolerate a relatively wide range of salinity are referred to as euryhaline organisms. This is possible because some fish have evolved **osmoregulatory** mechanisms to survive in all kinds of aquatic environments. When they live in fresh water, their

bodies tend to take up water because the environment is relatively hypotonic, as illustrated in [\[link\]](#)**a**. In such hypotonic environments, these fish do not drink much water. Instead, they pass a lot of very dilute urine, and they achieve electrolyte balance by active transport of salts through the gills. When they move to a hypertonic marine environment, these fish start drinking sea water; they excrete the excess salts through their gills and their urine, as illustrated in [\[link\]](#)**b**. Most marine invertebrates, on the other hand, may be isotonic with sea water (**osmoconformers**). Their body fluid concentrations conform to changes in seawater concentration. Cartilaginous fishes' salt composition of the blood is similar to bony fishes; however, the blood of sharks contains the organic compounds urea and trimethylamine oxide (TMAO). This does not mean that their electrolyte composition is similar to that of sea water. They achieve isotonicity with the sea by storing large concentrations of urea. These animals that secrete urea are called ureotelic animals. TMAO stabilizes proteins in the presence of high urea levels, preventing the disruption of peptide bonds that would occur in other animals exposed to similar levels of urea. Sharks are cartilaginous fish with a rectal gland to secrete salt and assist in osmoregulation.





(a) Osmoregulation in a freshwater environment



(b) Osmoregulation in a saltwater environment

Fish are osmoregulators, but must use different mechanisms to survive in (a) freshwater or (b) saltwater environments. (credit: modification of work by Duane Raver, NOAA)

### Note:

#### Career Connection

#### Dialysis Technician

Dialysis is a medical process of removing wastes and excess water from the blood by diffusion and ultrafiltration. When kidney function fails, dialysis must be done to artificially rid the body of wastes. This is a vital process to keep patients alive. In some cases, the patients undergo artificial dialysis until they are eligible for a kidney transplant. In others who are not candidates for kidney transplants, dialysis is a life-long necessity.

Dialysis technicians typically work in hospitals and clinics. While some roles in this field include equipment development and maintenance, most dialysis technicians work in direct patient care. Their on-the-job duties, which typically occur under the direct supervision of a registered nurse, focus on providing dialysis treatments. This can include reviewing patient history and current condition, assessing and responding to patient needs before and during treatment, and monitoring the dialysis process. Treatment may include taking and reporting a patient's vital signs and preparing solutions and equipment to ensure accurate and sterile procedures.

## **Section Summary**

Solute concentrations across a semi-permeable membranes influence the movement of water and solutes across the membrane. It is the number of solute molecules and not the molecular size that is important in osmosis. Osmoregulation and osmotic balance are important bodily functions, resulting in water and salt balance. Not all solutes can pass through a semi-permeable membrane. Osmosis is the movement of water across the membrane. Osmosis occurs to equalize the number of solute molecules across a semi-permeable membrane by the movement of water to the side of higher solute concentration. Facilitated diffusion utilizes protein channels to move solute molecules from areas of higher to lower concentration while active transport mechanisms are required to move solutes against concentration gradients. Osmolarity is measured in units of milliequivalents or milliosmoles, both of which take into consideration the number of solute particles and the charge on them. Fish that live in fresh water or saltwater adapt by being osmoregulators or osmoconformers.

## **Review Questions**

### **Exercise:**



**Problem:**

When a dehydrated human patient needs to be given fluids intravenously, he or she is given:

- a. water, which is hypotonic with respect to body fluids
- b. saline at a concentration that is isotonic with respect to body fluids
- c. glucose because it is a non-electrolyte
- d. blood

---

**Solution:**

B

**Exercise:**

**Problem:** The sodium ion is at the highest concentration in:

- a. intracellular fluid
- b. extracellular fluid
- c. blood plasma
- d. none of the above

---

**Solution:**

B

**Exercise:**

**Problem:** Cells in a hypertonic solution tend to:

- a. shrink due to water loss
- b. swell due to water gain
- c. stay the same size due to water moving into and out of the cell at the same rate

d. none of the above

---

**Solution:**

A

## Free Response

**Exercise:**

**Problem:**

Why is excretion important in order to achieve osmotic balance?

---

**Solution:**

Excretion allows an organism to rid itself of waste molecules that could be toxic if allowed to accumulate. It also allows the organism to keep the amount of water and dissolved solutes in balance.

**Exercise:**

**Problem:**

Why do electrolyte ions move across membranes by active transport?

---

**Solution:**

Electrolyte ions often require special mechanisms to cross the semi-permeable membranes in the body. Active transport is the movement against a concentration gradient.

## Glossary

electrolyte

solute that breaks down into ions when dissolved in water

molality

number of moles of solute per kilogram of solvent

molarity

number of moles of solute per liter of solution

mole

gram equivalent of the molecular weight of a substance

non-electrolyte

solute that does not break down into ions when dissolved in water

osmoconformer

organism that changes its tonicity based on its environment

osmoregulation

mechanism by which water and solute concentrations are maintained at desired levels

osmoregulator

organism that maintains its tonicity irrespective of its environment

osmotic balance

balance of the amount of water and salt input and output to and from a biological system without disturbing the desired osmotic pressure and solute concentration in every compartment

osmotic pressure

pressure exerted on a membrane to equalize solute concentration on either side

semi-permeable membrane

membrane that allows only certain solutes to pass through

## The Kidneys and Osmoregulatory Organs

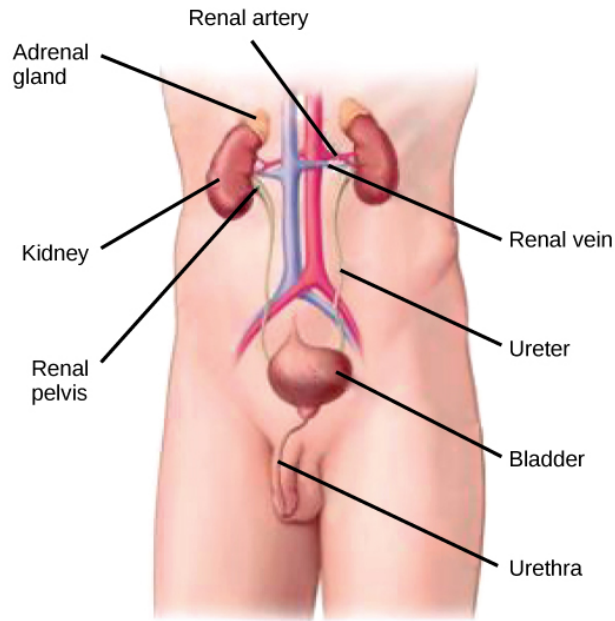
By the end of this section, you will be able to:

- Explain how the kidneys serve as the main osmoregulatory organs in mammalian systems
- Describe the structure of the kidneys and the functions of the parts of the kidney
- Describe how the nephron is the functional unit of the kidney and explain how it actively filters blood and generates urine
- Detail the three steps in the formation of urine: glomerular filtration, tubular reabsorption, and tubular secretion

Although the kidneys are the major osmoregulatory organ, the skin and lungs also play a role in the process. Water and electrolytes are lost through sweat glands in the skin, which helps moisturize and cool the skin surface, while the lungs expel a small amount of water in the form of mucous secretions and via evaporation of water vapor.

### Kidneys: The Main Osmoregulatory Organ

The **kidneys**, illustrated in [\[link\]](#), are a pair of bean-shaped structures that are located just below and posterior to the liver in the peritoneal cavity. The adrenal glands sit on top of each kidney and are also called the suprarenal glands. Kidneys filter blood and purify it. All the blood in the human body is filtered many times a day by the kidneys; these organs use up almost 25 percent of the oxygen absorbed through the lungs to perform this function. Oxygen allows the kidney cells to efficiently manufacture chemical energy in the form of ATP through aerobic respiration. The filtrate coming out of the kidneys is called **urine**.



Kidneys filter the blood, producing urine that is stored in the bladder prior to elimination through the urethra. (credit: modification of work by NCI)

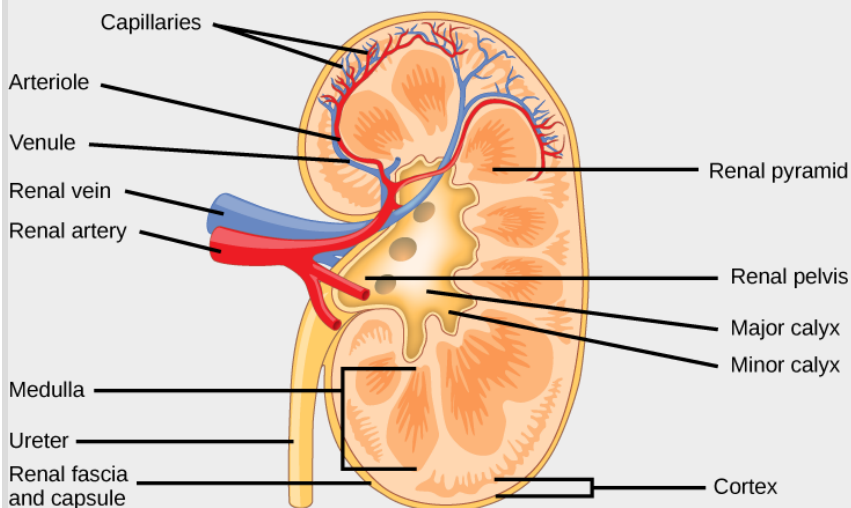
## Kidney Structure

Externally, the kidneys are surrounded by three layers, illustrated in [\[link\]](#). The outermost layer is a tough connective tissue layer called the **renal fascia**. The second layer is called the **perirenal fat capsule**, which helps anchor the kidneys in place. The third and innermost layer is the **renal capsule**. Internally, the kidney has three regions—an outer **cortex**, a **medulla** in the middle, and the **renal pelvis** in the region called the **hilum** of the kidney. The hilum is the concave part of the bean-shape where blood vessels and nerves enter and exit the kidney; it is also the point of exit for the ureters. The renal cortex is granular due to the presence of **nephrons**—the functional unit of the kidney. The medulla consists of multiple pyramidal tissue masses, called the **renal pyramids**. In between the pyramids are spaces called **renal columns** through which the blood vessels

pass. The tips of the pyramids, called renal papillae, point toward the renal pelvis. There are, on average, eight renal pyramids in each kidney. The renal pyramids along with the adjoining cortical region are called the **lobes of the kidney**. The renal pelvis leads to the **ureter** on the outside of the kidney. On the inside of the kidney, the renal pelvis branches out into two or three extensions called the major **calyces**, which further branch into the minor calyces. The ureters are urine-bearing tubes that exit the kidney and empty into the **urinary bladder**.

**Note:**

**Art Connection**



The internal structure of the kidney is shown.  
(credit: modification of work by NCI)

Which of the following statements about the kidney is false?

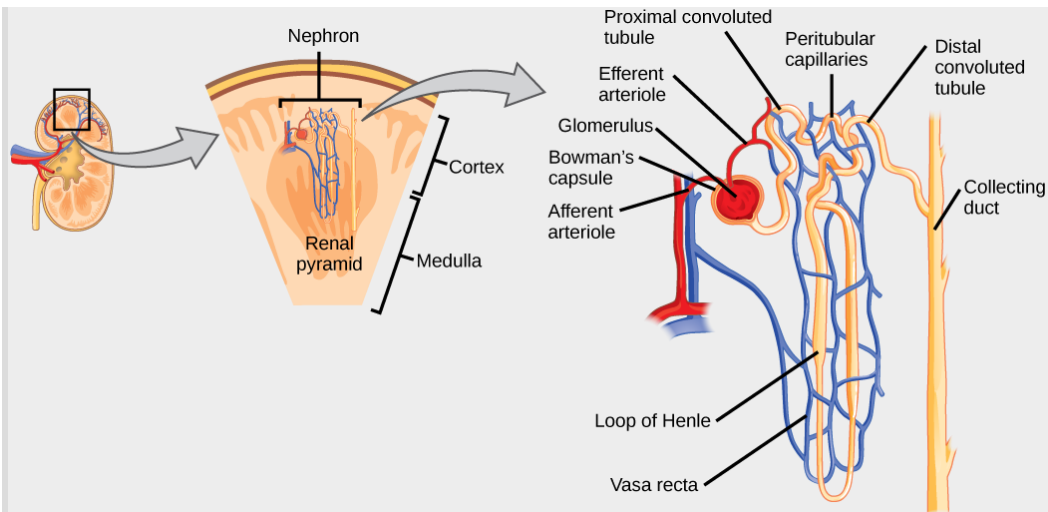
- a. The renal pelvis drains into the ureter.
- b. The renal pyramids are in the medulla.
- c. The cortex covers the capsule.
- d. Nephrons are in the renal cortex.

Because the kidney filters blood, its network of blood vessels is an important component of its structure and function. The arteries, veins, and nerves that supply the kidney enter and exit at the renal hilum. Renal blood supply starts with the branching of the aorta into the **renal arteries** (which are each named based on the region of the kidney they pass through) and ends with the exiting of the **renal veins** to join the **inferior vena cava**. The renal arteries split into several **segmental arteries** upon entering the kidneys. Each segmental artery splits further into several **interlobar arteries** and enters the renal columns, which supply the renal lobes. The interlobar arteries split at the junction of the renal cortex and medulla to form the **arcuate arteries**. The arcuate “bow shaped” arteries form arcs along the base of the medullary pyramids. **Cortical radiate arteries**, as the name suggests, radiate out from the arcuate arteries. The cortical radiate arteries branch into numerous afferent arterioles, and then enter the capillaries supplying the nephrons. Veins trace the path of the arteries and have similar names, except there are no segmental veins.

As mentioned previously, the functional unit of the kidney is the nephron, illustrated in [\[link\]](#). Each kidney is made up of over one million nephrons that dot the renal cortex, giving it a granular appearance when sectioned sagittally. There are two types of nephrons—**cortical nephrons** (85 percent), which are deep in the renal cortex, and **juxtamedullary nephrons** (15 percent), which lie in the renal cortex close to the renal medulla. A nephron consists of three parts—a **renal corpuscle**, a **renal tubule**, and the associated capillary network, which originates from the cortical radiate arteries.

**Note:**

Art Connection



The nephron is the functional unit of the kidney. The glomerulus and convoluted tubules are located in the kidney cortex, while collecting ducts are located in the pyramids of the medulla. (credit: modification of work by NIDDK)

Which of the following statements about the nephron is false?

- The collecting duct empties into the distal convoluted tubule.
- The Bowman's capsule surrounds the glomerulus.
- The loop of Henle is between the proximal and distal convoluted tubules.
- The loop of Henle empties into the distal convoluted tubule.

## Renal Corpuscle

The renal corpuscle, located in the renal cortex, is made up of a network of capillaries known as the **glomerulus** and the capsule, a cup-shaped chamber that surrounds it, called the glomerular or **Bowman's capsule**.



## Renal Tubule

The renal tubule is a long and convoluted structure that emerges from the glomerulus and can be divided into three parts based on function. The first part is called the **proximal convoluted tubule (PCT)** due to its proximity to the glomerulus; it stays in the renal cortex. The second part is called the **loop of Henle**, or nephritic loop, because it forms a loop (with **descending** and **ascending limbs**) that goes through the renal medulla. The third part of the renal tubule is called the **distal convoluted tubule (DCT)** and this part is also restricted to the renal cortex. The DCT, which is the last part of the nephron, connects and empties its contents into collecting ducts that line the medullary pyramids. The collecting ducts amass contents from multiple nephrons and fuse together as they enter the papillae of the renal medulla.

## Capillary Network within the Nephron

The capillary network that originates from the renal arteries supplies the nephron with blood that needs to be filtered. The branch that enters the glomerulus is called the **afferent arteriole**. The branch that exits the glomerulus is called the **efferent arteriole**. Within the glomerulus, the network of capillaries is called the glomerular capillary bed. Once the efferent arteriole exits the glomerulus, it forms the **peritubular capillary network**, which surrounds and interacts with parts of the renal tubule. In cortical nephrons, the peritubular capillary network surrounds the PCT and DCT. In juxtamedullary nephrons, the peritubular capillary network forms a network around the loop of Henle and is called the **vasa recta**.

### Note:

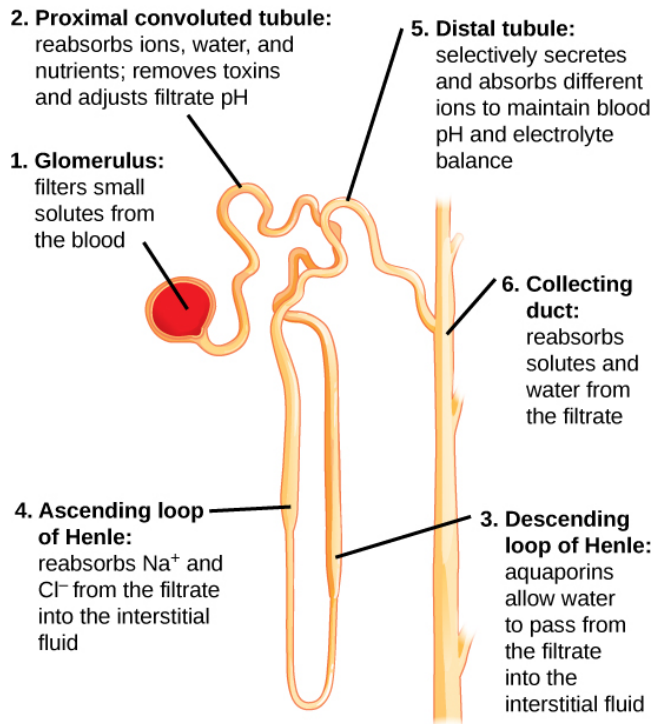
[Link to Learning](#)



Go to [this website](#) to see another coronal section of the kidney and to explore an animation of the workings of nephrons.

## Kidney Function and Physiology

Kidneys filter blood in a three-step process. First, the nephrons filter blood that runs through the capillary network in the glomerulus. Almost all solutes, except for proteins, are filtered out into the glomerulus by a process called **glomerular filtration**. Second, the filtrate is collected in the renal tubules. Most of the solutes get reabsorbed in the PCT by a process called **tubular reabsorption**. In the loop of Henle, the filtrate continues to exchange solutes and water with the renal medulla and the peritubular capillary network. Water is also reabsorbed during this step. Then, additional solutes and wastes are secreted into the kidney tubules during **tubular secretion**, which is, in essence, the opposite process to tubular reabsorption. The collecting ducts collect filtrate coming from the nephrons and fuse in the medullary papillae. From here, the papillae deliver the filtrate, now called urine, into the minor calyces that eventually connect to the ureters through the renal pelvis. This entire process is illustrated in [\[link\]](#).



Each part of the nephron performs a different function in filtering waste and maintaining homeostatic balance. (1) The glomerulus forces small solutes out of the blood by pressure. (2) The proximal convoluted tubule reabsorbs ions, water, and nutrients from the filtrate into the interstitial fluid, and actively transports toxins and drugs from the interstitial fluid into the filtrate. The proximal convoluted tubule also adjusts blood pH by selectively secreting ammonia ( $\text{NH}_3$ ) into the filtrate, where it reacts with  $\text{H}^+$  to form  $\text{NH}_4^+$ . The more acidic the filtrate, the more ammonia is secreted. (3) The descending loop of Henle is lined

with cells containing aquaporins that allow water to pass from the filtrate into the interstitial fluid.

(4) In the thin part of the ascending loop of Henle,  $\text{Na}^+$  and  $\text{Cl}^-$  ions diffuse into the interstitial fluid. In the thick part, these same ions are actively transported into the interstitial fluid. Because salt but not water is lost, the filtrate becomes more dilute as it travels up the limb. (5) In the distal convoluted tubule,  $\text{K}^+$  and  $\text{H}^+$  ions are selectively secreted into the filtrate, while  $\text{Na}^+$ ,  $\text{Cl}^-$ , and  $\text{HCO}_3^-$  ions are reabsorbed to maintain pH and electrolyte balance in the blood. (6) The collecting duct reabsorbs solutes and water from the filtrate, forming dilute urine. (credit: modification of work by NIDDK)

## **Glomerular Filtration**

Glomerular filtration filters out most of the solutes due to high blood pressure and specialized membranes in the afferent arteriole. The blood pressure in the glomerulus is maintained independent of factors that affect systemic blood pressure. The “leaky” connections between the endothelial cells of the glomerular capillary network allow solutes to pass through easily. All solutes in the glomerular capillaries, except for macromolecules like proteins, pass through by passive diffusion. There is no energy requirement at this stage of the filtration process. **Glomerular filtration**

**rate (GFR)** is the volume of glomerular filtrate formed per minute by the kidneys. GFR is regulated by multiple mechanisms and is an important indicator of kidney function.

**Note:**

Link to Learning



To learn more about the vascular system of kidneys, click through [this review](#) and the steps of blood flow.

## Tubular Reabsorption and Secretion

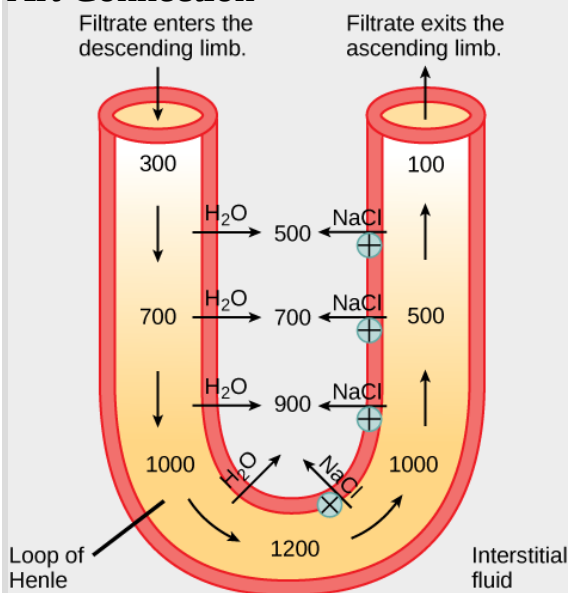
Tubular reabsorption occurs in the PCT part of the renal tubule. Almost all nutrients are reabsorbed, and this occurs either by passive or active transport. Reabsorption of water and some key electrolytes are regulated and can be influenced by hormones. Sodium ( $\text{Na}^+$ ) is the most abundant ion and most of it is reabsorbed by active transport and then transported to the peritubular capillaries. Because  $\text{Na}^+$  is actively transported out of the tubule, water follows it to even out the osmotic pressure. Water is also independently reabsorbed into the peritubular capillaries due to the presence of aquaporins, or water channels, in the PCT. This occurs due to the low blood pressure and high osmotic pressure in the peritubular capillaries. However, every solute has a **transport maximum** and the excess is not reabsorbed.

In the loop of Henle, the permeability of the membrane changes. The descending limb is permeable to water, not solutes; the opposite is true for the ascending limb. Additionally, the loop of Henle invades the renal

medulla, which is naturally high in salt concentration and tends to absorb water from the renal tubule and concentrate the filtrate. The osmotic gradient increases as it moves deeper into the medulla. Because two sides of the loop of Henle perform opposing functions, as illustrated in [\[link\]](#), it acts as a **countercurrent multiplier**. The vasa recta around it acts as the **countercurrent exchanger**.

### Note:

#### Art Connection



The loop of Henle acts as a countercurrent multiplier that uses energy to create concentration gradients. The descending limb is water permeable. Water flows from the filtrate to the interstitial fluid, so osmolality inside the limb increases as it descends into the renal medulla. At the bottom, the osmolality is higher inside the loop than in

the interstitial fluid. Thus, as filtrate enters the ascending limb,  $\text{Na}^+$  and  $\text{Cl}^-$  ions exit through ion channels present in the cell membrane. Further up,  $\text{Na}^+$  is actively transported out of the filtrate and  $\text{Cl}^-$  follows. Osmolarity is given in units of milliosmoles per liter (mOsm/L).

Loop diuretics are drugs sometimes used to treat hypertension. These drugs inhibit the reabsorption of  $\text{Na}^+$  and  $\text{Cl}^-$  ions by the ascending limb of the loop of Henle. A side effect is that they increase urination. Why do you think this is the case?

By the time the filtrate reaches the DCT, most of the urine and solutes have been reabsorbed. If the body requires additional water, all of it can be reabsorbed at this point. Further reabsorption is controlled by hormones, which will be discussed in a later section. Excretion of wastes occurs due to lack of reabsorption combined with tubular secretion. Undesirable products like metabolic wastes, urea, uric acid, and certain drugs, are excreted by tubular secretion. Most of the tubular secretion happens in the DCT, but some occurs in the early part of the collecting duct. Kidneys also maintain an acid-base balance by secreting excess  $\text{H}^+$  ions.

Although parts of the renal tubules are named proximal and distal, in a cross-section of the kidney, the tubules are placed close together and in contact with each other and the glomerulus. This allows for exchange of chemical messengers between the different cell types. For example, the DCT ascending limb of the loop of Henle has masses of cells called **macula densa**, which are in contact with cells of the afferent arterioles called **juxtaglomerular cells**. Together, the macula densa and juxtaglomerular

cells form the juxtaglomerular complex (JGC). The JGC is an endocrine structure that secretes the enzyme renin and the hormone erythropoietin. When hormones trigger the macula densa cells in the DCT due to variations in blood volume, blood pressure, or electrolyte balance, these cells can immediately communicate the problem to the capillaries in the afferent and efferent arterioles, which can constrict or relax to change the glomerular filtration rate of the kidneys.

**Note:****Career Connection****Nephrologist**

A nephrologist studies and deals with diseases of the kidneys—both those that cause kidney failure (such as diabetes) and the conditions that are produced by kidney disease (such as hypertension). Blood pressure, blood volume, and changes in electrolyte balance come under the purview of a nephrologist.

Nephrologists usually work with other physicians who refer patients to them or consult with them about specific diagnoses and treatment plans. Patients are usually referred to a nephrologist for symptoms such as blood or protein in the urine, very high blood pressure, kidney stones, or renal failure.

Nephrology is a subspecialty of internal medicine. To become a nephrologist, medical school is followed by additional training to become certified in internal medicine. An additional two or more years is spent specifically studying kidney disorders and their accompanying effects on the body.

**Section Summary**

The kidneys are the main osmoregulatory organs in mammalian systems; they function to filter blood and maintain the osmolarity of body fluids at 300 mOsm. They are surrounded by three layers and are made up internally of three distinct regions—the cortex, medulla, and pelvis.



The blood vessels that transport blood into and out of the kidneys arise from and merge with the aorta and inferior vena cava, respectively. The renal arteries branch out from the aorta and enter the kidney where they further divide into segmental, interlobar, arcuate, and cortical radiate arteries.

The nephron is the functional unit of the kidney, which actively filters blood and generates urine. The nephron is made up of the renal corpuscle and renal tubule. Cortical nephrons are found in the renal cortex, while juxtamedullary nephrons are found in the renal cortex close to the renal medulla. The nephron filters and exchanges water and solutes with two sets of blood vessels and the tissue fluid in the kidneys.

There are three steps in the formation of urine: glomerular filtration, which occurs in the glomerulus; tubular reabsorption, which occurs in the renal tubules; and tubular secretion, which also occurs in the renal tubules.

## Art Connections

### Exercise:

#### Problem:

[\[link\]](#) Which of the following statements about the kidney is false?

- a. The renal pelvis drains into the ureter.
- b. The renal pyramids are in the medulla.
- c. The cortex covers the capsule.
- d. Nephrons are in the renal cortex.

---

#### Solution:

[\[link\]](#) C

### Exercise:

#### Problem:

[\[link\]](#) Which of the following statements about the nephron is false?

- a. The collecting duct empties into the distal convoluted tubule.
- b. The Bowman's capsule surrounds the glomerulus.
- c. The loop of Henle is between the proximal and distal convoluted tubules.
- d. The loop of Henle empties into the distal convoluted tubule.

---

**Solution:**

[\[link\]](#) A

**Exercise:**

**Problem:**

[\[link\]](#) Loop diuretics are drugs sometimes used to treat hypertension. These drugs inhibit the reabsorption of  $\text{Na}^+$  and  $\text{Cl}^-$  ions by the ascending limb of the loop of Henle. A side effect is that they increase urination. Why do you think this is the case?

---

**Solution:**

[\[link\]](#) Loop diuretics decrease the excretion of salt into the renal medulla, thereby reducing its osmolality. As a result, less water is excreted into the medulla by the descending limb, and more water is excreted as urine.

## Review Questions

**Exercise:**

**Problem:** The macula densa is/are:

- a. present in the renal medulla.
- b. dense tissue present in the outer layer of the kidney.
- c. cells present in the DCT and collecting tubules.
- d. present in blood capillaries.

---

**Solution:**

C

**Exercise:**

**Problem:** The osmolarity of body fluids is maintained at \_\_\_\_\_.

- a. 100 mOsm
- b. 300 mOsm
- c. 1000 mOsm
- d. it is not constantly maintained

---

**Solution:**

B

**Exercise:**

**Problem:**

The gland located at the top of the kidney is the \_\_\_\_\_ gland.

- a. adrenal
- b. pituitary
- c. thyroid
- d. thymus

---

**Solution:**

A

**Free Response**

**Exercise:**

**Problem:**

Why are the loop of Henle and vasa recta important for the formation of concentrated urine?

---

**Solution:**

The loop of Henle is part of the renal tubule that loops into the renal medulla. In the loop of Henle, the filtrate exchanges solutes and water with the renal medulla and the vasa recta (the peritubular capillary network). The vasa recta acts as the countercurrent exchanger. The kidneys maintain the osmolality of the rest of the body at a constant 300 mOsm by concentrating the filtrate as it passes through the loop of Henle.

**Exercise:**

**Problem:** Describe the structure of the kidney.

---

**Solution:**

Externally, the kidneys are surrounded by three layers. The outermost layer is a tough connective tissue layer called the renal fascia. The second layer is called the perirenal fat capsule, which helps anchor the kidneys in place. The third and innermost layer is the renal capsule. Internally, the kidney has three regions—an outer cortex, a medulla in the middle, and the renal pelvis in the region called the hilum of the kidney, which is the concave part of the “bean” shape.

**Glossary**

afferent arteriole

arteriole that branches from the cortical radiate artery and enters the glomerulus

arcuate artery

artery that branches from the interlobar artery and arches over the base of the renal pyramids

ascending limb

part of the loop of Henle that ascends from the renal medulla to the renal cortex

Bowman's capsule

structure that encloses the glomerulus

calyx

structure that connects the renal pelvis to the renal medulla

cortex (animal)

outer layer of an organ like the kidney or adrenal gland

cortical nephron

nephron that lies in the renal cortex

cortical radiate artery

artery that radiates from the arcuate arteries into the renal cortex

countercurrent exchanger

peritubular capillary network that allows exchange of solutes and water from the renal tubules

countercurrent multiplier

osmotic gradient in the renal medulla that is responsible for concentration of urine

descending limb

part of the loop of Henle that descends from the renal cortex into the renal medulla

distal convoluted tubule (DCT)

part of the renal tubule that is the most distant from the glomerulus

efferent arteriole

arteriole that exits from the glomerulus

glomerular filtration

filtration of blood in the glomerular capillary network into the glomerulus

glomerular filtration rate (GFR)

amount of filtrate formed by the glomerulus per minute

glomerulus (renal)

part of the renal corpuscle that contains the capillary network

hilum

region in the renal pelvis where blood vessels, nerves, and ureters bunch before entering or exiting the kidney

inferior vena cava

one of the main veins in the human body

interlobar artery

artery that branches from the segmental artery and travels in between the renal lobes

juxtaglomerular cell

cell in the afferent and efferent arterioles that responds to stimuli from the macula densa

juxtamedullary nephron

nephron that lies in the cortex but close to the renal medulla

kidney

organ that performs excretory and osmoregulatory functions

lobes of the kidney

renal pyramid along with the adjoining cortical region

loop of Henle

part of the renal tubule that loops into the renal medulla

macula densa

group of cells that senses changes in sodium ion concentration; present in parts of the renal tubule and collecting ducts

medulla

middle layer of an organ like the kidney or adrenal gland

nephron

functional unit of the kidney

perirenal fat capsule

fat layer that suspends the kidneys

peritubular capillary network

capillary network that surrounds the renal tubule after the efferent artery exits the glomerulus

proximal convoluted tubule (PCT)

part of the renal tubule that lies close to the glomerulus

renal artery

branch of the artery that enters the kidney

renal capsule

layer that encapsulates the kidneys

renal column

area of the kidney through which the interlobar arteries travel in the process of supplying blood to the renal lobes

renal corpuscle

glomerulus and the Bowman's capsule together

renal fascia

connective tissue that supports the kidneys

renal pelvis

region in the kidney where the calyces join the ureters

renal pyramid

conical structure in the renal medulla

renal tubule

tubule of the nephron that arises from the glomerulus

renal vein

branch of a vein that exits the kidney and joins the inferior vena cava

segmental artery

artery that branches from the renal artery

transport maximum

maximum amount of solute that can be transported out of the renal tubules during reabsorption

tubular reabsorption

reclamation of water and solutes that got filtered out in the glomerulus

tubular secretion

process of secretion of wastes that do not get reabsorbed

ureter

urine-bearing tube coming out of the kidney; carries urine to the bladder

urinary bladder

structure that the ureters empty the urine into; stores urine

urine

filtrate produced by kidneys that gets excreted out of the body

vasa recta

peritubular network that surrounds the loop of Henle of the juxtamedullary nephrons



## Excretion Systems

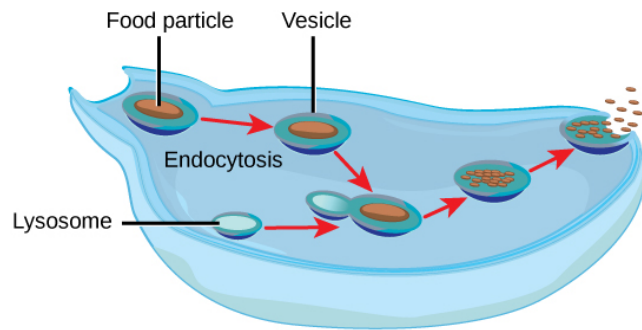
By the end of this section, you will be able to:

- Explain how vacuoles, present in microorganisms, work to excrete waste
- Describe the way in which flame cells and nephridia in worms perform excretory functions and maintain osmotic balance
- Explain how insects use Malpighian tubules to excrete wastes and maintain osmotic balance

Microorganisms and invertebrate animals use more primitive and simple mechanisms to get rid of their metabolic wastes than the mammalian system of kidney and urinary function. Three excretory systems evolved in organisms before complex kidneys: vacuoles, flame cells, and Malpighian tubules.

### **Contractile Vacuoles in Microorganisms**

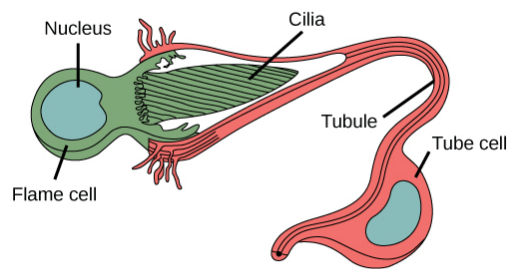
The most fundamental feature of life is the presence of a cell. In other words, a cell is the simplest functional unit of a life. Bacteria are unicellular, prokaryotic organisms that have some of the least complex life processes in place; however, prokaryotes such as bacteria do not contain membrane-bound vacuoles. The cells of microorganisms like bacteria, protozoa, and fungi are bound by cell membranes and use them to interact with the environment. Some cells, including some leucocytes in humans, are able to engulf food by endocytosis—the formation of vesicles by involution of the cell membrane within the cells. The same vesicles are able to interact and exchange metabolites with the intracellular environment. In some unicellular eukaryotic organisms such as the amoeba, shown in [\[link\]](#), cellular wastes and excess water are excreted by exocytosis, when the contractile vacuoles merge with the cell membrane and expel wastes into the environment. Contractile vacuoles (CV) should not be confused with vacuoles, which store food or water.



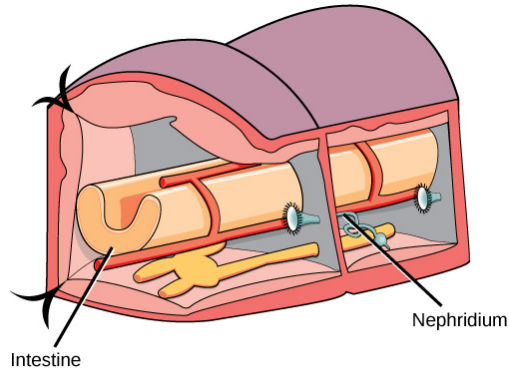
Some unicellular organisms, such as the amoeba, ingest food by endocytosis. The food vesicle fuses with a lysosome, which digests the food. Waste is excreted by exocytosis.

## Flame Cells of Planaria and Nephridia of Worms

As multi-cellular systems evolved to have organ systems that divided the metabolic needs of the body, individual organs evolved to perform the excretory function. Planaria are flatworms that live in fresh water. Their excretory system consists of two tubules connected to a highly branched duct system. The cells in the tubules are called **flame cells** (or **protonephridia**) because they have a cluster of cilia that looks like a flickering flame when viewed under the microscope, as illustrated in [\[link\]](#)<sup>a</sup>. The cilia propel waste matter down the tubules and out of the body through excretory pores that open on the body surface; cilia also draw water from the interstitial fluid, allowing for filtration. Any valuable metabolites are recovered by reabsorption. Flame cells are found in flatworms, including parasitic tapeworms and free-living planaria. They also maintain the organism's osmotic balance.



(a) Flame cell of a planarian



(b) Nephridium of an earthworm

In the excretory system of the (a) planaria, cilia of flame cells propel waste through a tubule formed by a tube cell. Tubules are connected into branched structures that lead to pores located all along the sides of the body. The filtrate is secreted through these pores. In (b) annelids such as earthworms, nephridia filter fluid from the coelom, or body cavity. Beating cilia at the opening of the nephridium draw water from the coelom into a tubule. As the filtrate passes down the tubules, nutrients and other solutes are reabsorbed by capillaries. Filtered fluid containing nitrogenous and other wastes is stored in a bladder and then secreted through a pore in the side of the body.

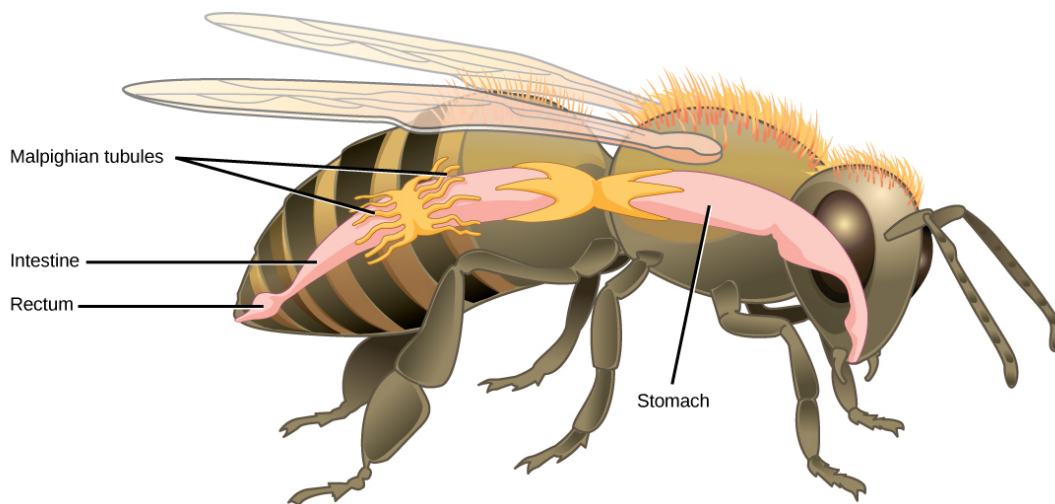
Earthworms (annelids) have slightly more evolved excretory structures called **nephridia**, illustrated in [\[link\]](#) **b**. A pair of nephridia is present on each segment of the earthworm. They are similar to flame cells in that they have a tubule with cilia. Excretion occurs through a pore called the **nephridiopore**. They are more evolved than the flame cells in that they have a system for tubular reabsorption by a capillary network before excretion.

## Malpighian Tubules of Insects

**Malpighian tubules** are found lining the gut of some species of arthropods, such as the bee illustrated in [\[link\]](#). They are usually found in pairs and the

number of tubules varies with the species of insect. Malpighian tubules are convoluted, which increases their surface area, and they are lined with **microvilli** for reabsorption and maintenance of osmotic balance.

Malpighian tubules work cooperatively with specialized glands in the wall of the rectum. Body fluids are not filtered as in the case of nephridia; urine is produced by tubular secretion mechanisms by the cells lining the Malpighian tubules that are bathed in hemolymph (a mixture of blood and interstitial fluid that is found in insects and other arthropods as well as most mollusks). Metabolic wastes like uric acid freely diffuse into the tubules. There are exchange pumps lining the tubules, which actively transport  $H^+$  ions into the cell and  $K^+$  or  $Na^+$  ions out; water passively follows to form urine. The secretion of ions alters the osmotic pressure which draws water, electrolytes, and nitrogenous waste (uric acid) into the tubules. Water and electrolytes are reabsorbed when these organisms are faced with low-water environments, and uric acid is excreted as a thick paste or powder. Not dissolving wastes in water helps these organisms to conserve water; this is especially important for life in dry environments.



Malpighian tubules of insects and other terrestrial arthropods remove nitrogenous wastes and other solutes from the hemolymph.  $Na^+$  and/or  $K^+$  ions are actively transported into the lumen of the tubules. Water then enters the tubules via osmosis, forming urine. The urine passes through the intestine,

and into the rectum. There, nutrients diffuse back into the hemolymph.  $\text{Na}^+$  and/or  $\text{K}^+$  ions are pumped into the hemolymph, and water follows. The concentrated waste is then excreted.

**Note:**

Link to Learning



See a dissected cockroach, including a close-up look at its Malpighian tubules, in this [video](#).

## Section Summary

Many systems have evolved for excreting wastes that are simpler than the kidney and urinary systems of vertebrate animals. The simplest system is that of contractile vacuoles present in microorganisms. Flame cells and nephridia in worms perform excretory functions and maintain osmotic balance. Some insects have evolved Malpighian tubules to excrete wastes and maintain osmotic balance.

## Review Questions

### Exercise:

**Problem:** Active transport of  $\text{K}^+$  in Malpighian tubules ensures that:

- a. water follows  $K^+$  to make urine
  - b. osmotic balance is maintained between waste matter and bodily fluids
  - c. both a and b
  - d. neither a nor b
- 

**Solution:**

C

**Exercise:**

**Problem:** Contractile vacuoles in microorganisms:

- a. exclusively perform an excretory function
  - b. can perform many functions, one of which is excretion of metabolic wastes
  - c. originate from the cell membrane
  - d. both b and c
- 

**Solution:**

D

**Exercise:**

**Problem:**

Flame cells are primitive excretory organs found in \_\_\_\_\_.

- a. arthropods
  - b. annelids
  - c. mammals
  - d. flatworms
- 

**Solution:**

D

## Free Response

### Exercise:

#### Problem:

Why might specialized organs have evolved for excretion of wastes?

---

#### Solution:

The removal of wastes, which could otherwise be toxic to an organism, is extremely important for survival. Having organs that specialize in this process and that operate separately from other organs provides a measure of safety for the organism.

### Exercise:

#### Problem:

Explain two different excretory systems other than the kidneys.

---

#### Solution:

(1) Microorganisms engulf food by endocytosis—the formation of vacuoles by involution of the cell membrane within the cells. The same vacuoles interact and exchange metabolites with the intracellular environment. Cellular wastes are excreted by exocytosis when the vacuoles merge with the cell membrane and excrete wastes into the environment. (2) Flatworms have an excretory system that consists of two tubules. The cells in the tubules are called flame cells; they have a cluster of cilia that propel waste matter down the tubules and out of the body. (3) Annelids have nephridia which have a tubule with cilia. Excretion occurs through a pore called the nephridiopore. Annelids have a system for tubular reabsorption by a capillary network before excretion. (4) Malpighian tubules are found in some species of arthropods. They are usually found in pairs, and the number of tubules varies with the species of insect. Malpighian tubules are convoluted,

which increases their surface area, and they are lined with microvilli for reabsorption and maintenance of osmotic balance. Metabolic wastes like uric acid freely diffuse into the tubules. Potassium ion pumps line the tubules, which actively transport out  $K^+$  ions, and water follows to form urine. Water and electrolytes are reabsorbed when these organisms are faced with low-water environments, and uric acid is excreted as a thick paste or powder. By not dissolving wastes in water, these organisms conserve water.

## **Glossary**

flame cell

(also, protonephridia) excretory cell found in flatworms

Malpighian tubule

excretory tubules found in arthropods

microvilli

cellular processes that increase the surface area of cells

nephridia

excretory structures found in annelids

nephridiopore

pore found at the end of nephridia



## Nitrogenous Wastes

By the end of this section, you will be able to:

- Compare and contrast the way in which aquatic animals and terrestrial animals can eliminate toxic ammonia from their systems
- Compare the major byproduct of ammonia metabolism in vertebrate animals to that of birds, insects, and reptiles

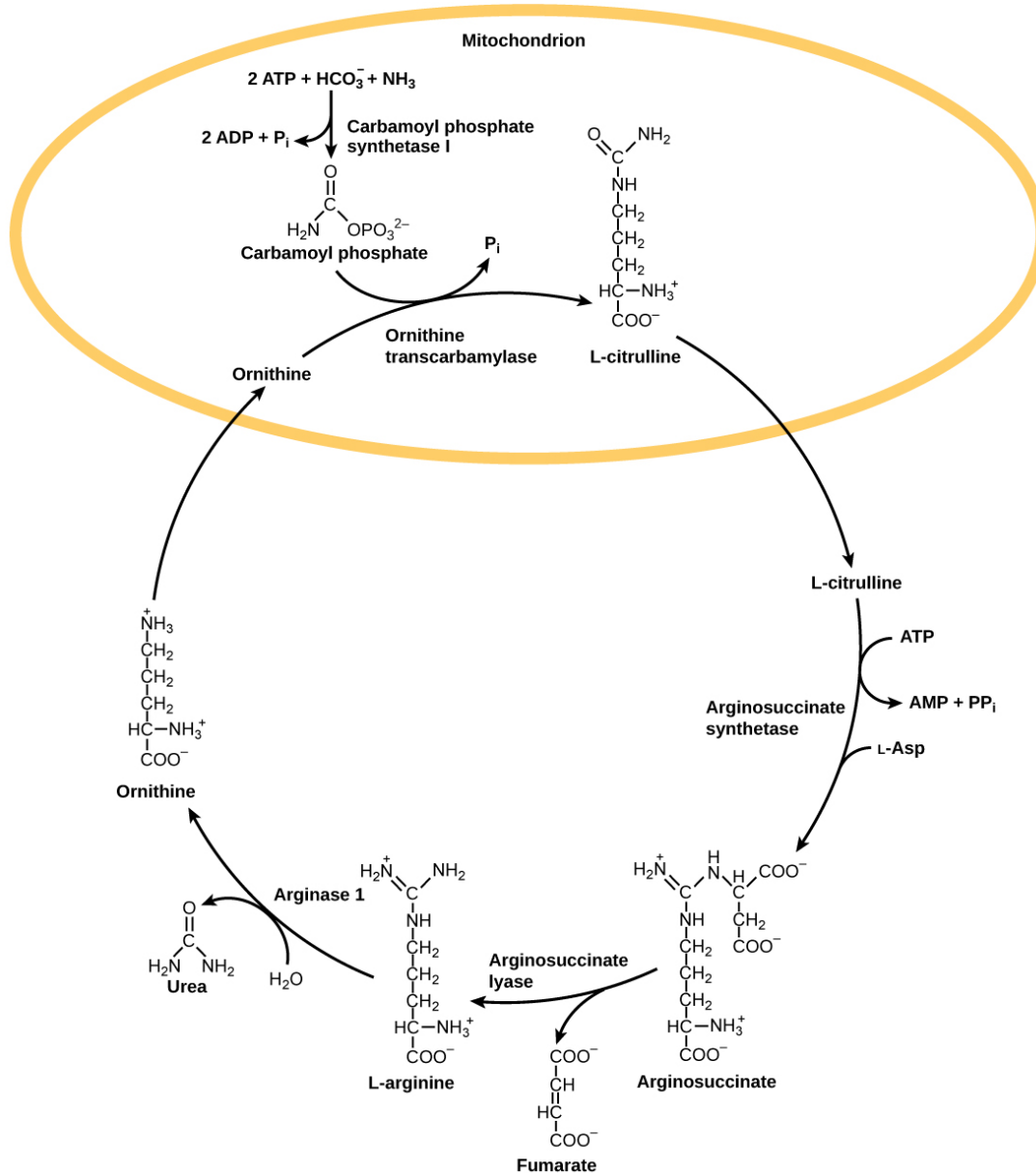
Of the four major macromolecules in biological systems, both proteins and nucleic acids contain nitrogen. During the catabolism, or breakdown, of nitrogen-containing macromolecules, carbon, hydrogen, and oxygen are extracted and stored in the form of carbohydrates and fats. Excess nitrogen is excreted from the body. Nitrogenous wastes tend to form toxic **ammonia**, which raises the pH of body fluids. The formation of ammonia itself requires energy in the form of ATP and large quantities of water to dilute it out of a biological system. Animals that live in aquatic environments tend to release ammonia into the water. Animals that excrete ammonia are said to be **ammonotelic**. Terrestrial organisms have evolved other mechanisms to excrete nitrogenous wastes. The animals must detoxify ammonia by converting it into a relatively nontoxic form such as urea or uric acid. Mammals, including humans, produce urea, whereas reptiles and many terrestrial invertebrates produce uric acid. Animals that secrete urea as the primary nitrogenous waste material are called **ureotelic** animals.

## Nitrogenous Waste in Terrestrial Animals: The Urea Cycle

The **urea cycle** is the primary mechanism by which mammals convert ammonia to urea. Urea is made in the liver and excreted in urine. The overall chemical reaction by which ammonia is converted to urea is  $2 \text{NH}_3$  (ammonia) +  $\text{CO}_2$  + 3 ATP +  $\text{H}_2\text{O}$  →  $\text{H}_2\text{N-CO-NH}_2$  (urea) + 2 ADP + 4  $\text{P}_i$  + AMP.

The urea cycle utilizes five intermediate steps, catalyzed by five different enzymes, to convert ammonia to urea, as shown in [\[link\]](#). The amino acid L-ornithine gets converted into different intermediates before being regenerated at the end of the urea cycle. Hence, the urea cycle is also referred to as the ornithine cycle. The enzyme ornithine transcarbamylase

catalyzes a key step in the urea cycle and its deficiency can lead to accumulation of toxic levels of ammonia in the body. The first two reactions occur in the mitochondria and the last three reactions occur in the cytosol. Urea concentration in the blood, called **blood urea nitrogen** or BUN, is used as an indicator of kidney function.






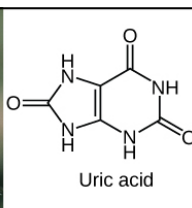
The urea cycle converts ammonia to urea.

**Note:****Evolution Connection****Excretion of Nitrogenous Waste**

The theory of evolution proposes that life started in an aquatic environment. It is not surprising to see that biochemical pathways like the urea cycle evolved to adapt to a changing environment when terrestrial life forms evolved. Arid conditions probably led to the evolution of the uric acid pathway as a means of conserving water.

**Nitrogenous Waste in Birds and Reptiles: Uric Acid**

Birds, reptiles, and most terrestrial arthropods convert toxic ammonia to **uric acid** or the closely related compound guanine (guano) instead of urea. Mammals also form some uric acid during breakdown of nucleic acids. Uric acid is a compound similar to purines found in nucleic acids. It is water insoluble and tends to form a white paste or powder; it is excreted by birds, insects, and reptiles. Conversion of ammonia to uric acid requires more energy and is much more complex than conversion of ammonia to urea [\[link\]](#).

	$\text{NH}_3$ Ammonia
	$\text{H}_2\text{N}-\text{C}(=\text{O})-\text{NH}_2$ Urea
	 Uric acid
(a) Many invertebrates and aquatic species excrete ammonia.	
(b) Mammals, many adult amphibians, and some marine species excrete urea.	
(c) Insects, land snails, birds, and many reptiles excrete uric acid.	

Nitrogenous waste is excreted in different forms by different species. These include (a) ammonia, (b) urea, and (c) uric acid.  
(credit a: modification of work by Eric Engbretson, USFWS;

credit b: modification of work by B. "Moose" Peterson, USFWS;  
credit c: modification of work by Dave Menke, USFWS)

**Note:**

**Everyday Connection**

**Gout**

Mammals use uric acid crystals as an **antioxidant** in their cells. However, too much uric acid tends to form kidney stones and may also cause a painful condition called gout, where uric acid crystals accumulate in the joints, as illustrated in [\[link\]](#). Food choices that reduce the amount of nitrogenous bases in the diet help reduce the risk of gout. For example, tea, coffee, and chocolate have purine-like compounds, called xanthines, and should be avoided by people with gout and kidney stones.



Gout causes the inflammation visible in this person's left big toe joint.  
(credit: "Gonzosft"/Wikimedia Commons)

## Section Summary

Ammonia is the waste produced by metabolism of nitrogen-containing compounds like proteins and nucleic acids. While aquatic animals can easily excrete ammonia into their watery surroundings, terrestrial animals have evolved special mechanisms to eliminate the toxic ammonia from their systems. Urea is the major byproduct of ammonia metabolism in vertebrate animals. Uric acid is the major byproduct of ammonia metabolism in birds, terrestrial arthropods, and reptiles.

## Review Questions

### Exercise:

**Problem:** BUN is \_\_\_\_\_.

- a. blood urea nitrogen
- b. blood uric acid nitrogen
- c. an indicator of blood volume
- d. an indicator of blood pressure

---

### Solution:

A

### Exercise:

#### Problem:

Human beings accumulate \_\_\_\_\_ before excreting nitrogenous waste.

- a. nitrogen
  - b. ammonia
  - c. urea
  - d. uric acid
-

**Solution:**

C

**Free Response****Exercise:****Problem:**

In terms of evolution, why might the urea cycle have evolved in organisms?

---

**Solution:**

It is believed that the urea cycle evolved to adapt to a changing environment when terrestrial life forms evolved. Arid conditions probably led to the evolution of the uric acid pathway as a means of conserving water.

**Exercise:**

**Problem:** Compare and contrast the formation of urea and uric acid.

---

**Solution:**

The urea cycle is the primary mechanism by which mammals convert ammonia to urea. Urea is made in the liver and excreted in urine. The urea cycle utilizes five intermediate steps, catalyzed by five different enzymes, to convert ammonia to urea. Birds, reptiles, and insects, on the other hand, convert toxic ammonia to uric acid instead of urea. Conversion of ammonia to uric acid requires more energy and is much more complex than conversion of ammonia to urea.

**Glossary**

ammonia

compound made of one nitrogen atom and three hydrogen atoms

ammonotelic

describes an animal that excretes ammonia as the primary waste material

antioxidant

agent that prevents cell destruction by reactive oxygen species

blood urea nitrogen (BUN)

estimate of urea in the blood and an indicator of kidney function

urea cycle

pathway by which ammonia is converted to urea

ureotelic

describes animals that secrete urea as the primary nitrogenous waste material

uric acid

byproduct of ammonia metabolism in birds, insects, and reptiles

Hormonal Control of Osmoregulatory Functions

By the end of this section, you will be able to:

- Explain how hormonal cues help the kidneys synchronize the osmotic needs of the body
- Describe how hormones like epinephrine, norepinephrine, renin-angiotensin, aldosterone, anti-diuretic hormone, and atrial natriuretic peptide help regulate waste elimination, maintain correct osmolarity, and perform other osmoregulatory functions

While the kidneys operate to maintain osmotic balance and blood pressure in the body, they also act in concert with hormones. Hormones are small molecules that act as messengers within the body. Hormones are typically secreted from one cell and travel in the bloodstream to affect a target cell in another portion of the body. Different regions of the nephron bear specialized cells that have receptors to respond to chemical messengers and hormones. [\[link\]](#) summarizes the hormones that control the osmoregulatory functions.

Hormones That Affect Osmoregulation		
Hormone	Where produced	Function
Epinephrine and Norepinephrine	Adrenal medulla	Can decrease kidney function temporarily by vasoconstriction
Renin	Kidney nephrons	Increases blood pressure by acting on angiotensinogen



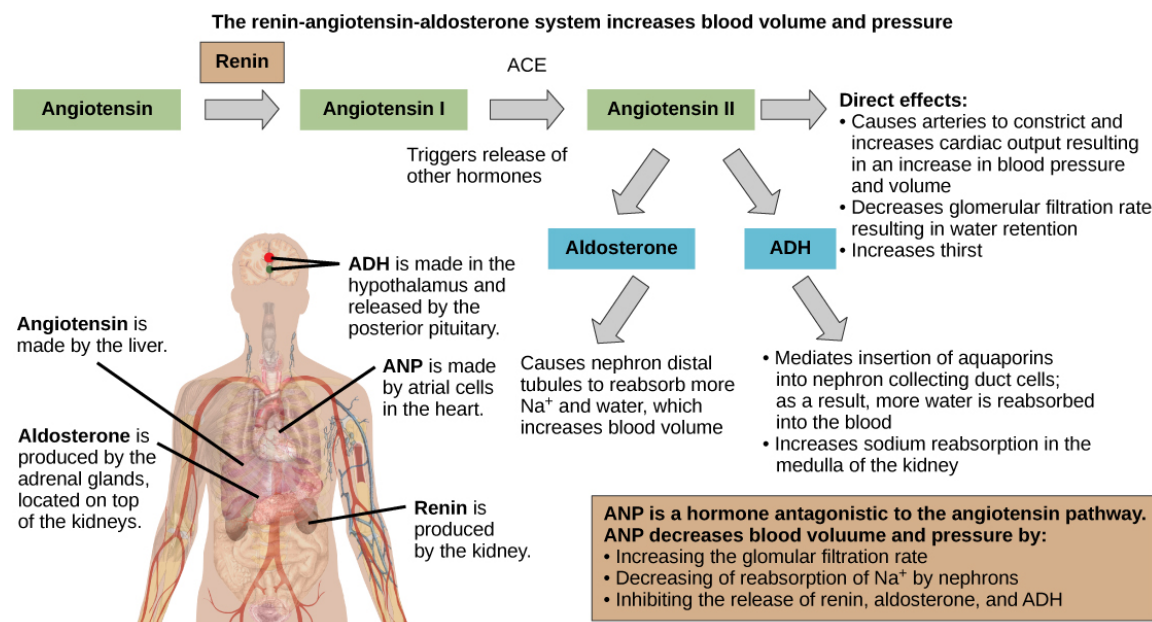
<b>Hormones That Affect Osmoregulation</b>		
<b>Hormone</b>	<b>Where produced</b>	<b>Function</b>
Angiotensin	Liver	Angiotensin II affects multiple processes and increases blood pressure
Aldosterone	Adrenal cortex	Prevents loss of sodium and water
Anti-diuretic hormone (vasopressin)	Hypothalamus (stored in the posterior pituitary)	Prevents water loss
Atrial natriuretic peptide	Heart atrium	Decreases blood pressure by acting as a vasodilator and increasing glomerular filtration rate; decreases sodium reabsorption in kidneys

## Epinephrine and Norepinephrine

Epinephrine and norepinephrine are released by the adrenal medulla and nervous system respectively. They are the flight/fight hormones that are released when the body is under extreme stress. During stress, much of the body's energy is used to combat imminent danger. Kidney function is halted temporarily by epinephrine and norepinephrine. These hormones function by acting directly on the smooth muscles of blood vessels to constrict them. Once the afferent arterioles are constricted, blood flow into the nephrons stops. These hormones go one step further and trigger the **renin-angiotensin-aldosterone** system.

## Renin-Angiotensin-Aldosterone

The renin-angiotensin-aldosterone system, illustrated in [\[link\]](#) proceeds through several steps to produce **angiotensin II**, which acts to stabilize blood pressure and volume. Renin (secreted by a part of the juxtaglomerular complex) is produced by the granular cells of the afferent and efferent arterioles. Thus, the kidneys control blood pressure and volume directly. Renin acts on angiotensinogen, which is made in the liver and converts it to **angiotensin I**. **Angiotensin converting enzyme (ACE)** converts angiotensin I to angiotensin II. Angiotensin II raises blood pressure by constricting blood vessels. It also triggers the release of the mineralocorticoid aldosterone from the adrenal cortex, which in turn stimulates the renal tubules to reabsorb more sodium. Angiotensin II also triggers the release of **anti-diuretic hormone (ADH)** from the hypothalamus, leading to water retention in the kidneys. It acts directly on the nephrons and decreases glomerular filtration rate. Medically, blood pressure can be controlled by drugs that inhibit ACE (called ACE inhibitors).



The renin-angiotensin-aldosterone system increases blood pressure and volume. The hormone ANP has antagonistic effects.

(credit: modification of work by Mikael Häggström)

## Mineralocorticoids

Mineralocorticoids are hormones synthesized by the adrenal cortex that affect osmotic balance. Aldosterone is a mineralocorticoid that regulates sodium levels in the blood. Almost all of the sodium in the blood is reclaimed by the renal tubules under the influence of aldosterone. Because sodium is always reabsorbed by active transport and water follows sodium to maintain osmotic balance, aldosterone manages not only sodium levels but also the water levels in body fluids. In contrast, the aldosterone also stimulates potassium secretion concurrently with sodium reabsorption. In contrast, absence of aldosterone means that no sodium gets reabsorbed in the renal tubules and all of it gets excreted in the urine. In addition, the daily dietary potassium load is not secreted and the retention of  $K^+$  can cause a dangerous increase in plasma  $K^+$  concentration. Patients who have Addison's disease have a failing adrenal cortex and cannot produce aldosterone. They lose sodium in their urine constantly, and if the supply is not replenished, the consequences can be fatal.

## Antidiurectic Hormone

As previously discussed, antidiuretic hormone or ADH (also called **vasopressin**), as the name suggests, helps the body conserve water when body fluid volume, especially that of blood, is low. It is formed by the hypothalamus and is stored and released from the posterior pituitary. It acts by inserting aquaporins in the collecting ducts and promotes reabsorption of water. ADH also acts as a vasoconstrictor and increases blood pressure during hemorrhaging.

## Atrial Natriuretic Peptide Hormone

The atrial natriuretic peptide (ANP) lowers blood pressure by acting as a **vasodilator**. It is released by cells in the atrium of the heart in response to

high blood pressure and in patients with sleep apnea. ANP affects salt release, and because water passively follows salt to maintain osmotic balance, it also has a diuretic effect. ANP also prevents sodium reabsorption by the renal tubules, decreasing water reabsorption (thus acting as a diuretic) and lowering blood pressure. Its actions suppress the actions of aldosterone, ADH, and renin.

## Section Summary

Hormonal cues help the kidneys synchronize the osmotic needs of the body. Hormones like epinephrine, norepinephrine, renin-angiotensin, aldosterone, anti-diuretic hormone, and atrial natriuretic peptide help regulate the needs of the body as well as the communication between the different organ systems.

## Review Questions

### Exercise:

**Problem:** Renin is made by \_\_\_\_\_.

- a. granular cells of the juxtaglomerular apparatus
- b. the kidneys
- c. the nephrons
- d. All of the above.

---

### Solution:

A

### Exercise:

**Problem:** Patients with Addison's disease \_\_\_\_\_.

- a. retain water
- b. retain salts

- c. lose salts and water
  - d. have too much aldosterone
- 

**Solution:**

C

**Exercise:**

**Problem:** Which hormone elicits the “fight or flight” response?

- a. epinephrine
  - b. mineralcorticoids
  - c. anti-diuretic hormone
  - d. thyroxine
- 

**Solution:**

A

**Free Response**

**Exercise:**

**Problem:**

Describe how hormones regulate blood pressure, blood volume, and kidney function.

---

**Solution:**

Hormones are small molecules that act as messengers within the body. Different regions of the nephron bear specialized cells, which have receptors to respond to chemical messengers and hormones. The hormones carry messages to the kidney. These hormonal cues help the kidneys synchronize the osmotic needs of the body. Hormones like

epinephrine, norepinephrine, renin-angiotensin, aldosterone, anti-diuretic hormone, and atrial natriuretic peptide help regulate the needs of the body as well as the communication between the different organ systems.

### **Exercise:**

#### **Problem:**

How does the renin-angiotensin-aldosterone mechanism function?  
Why is it controlled by the kidneys?

---

#### **Solution:**

The renin-angiotensin-aldosterone system acts through several steps to produce angiotensin II, which acts to stabilize blood pressure and volume. Thus, the kidneys control blood pressure and volume directly. Renin acts on angiotensinogen, which is made in the liver and converts it to angiotensin I. ACE (angiotensin converting enzyme) converts angiotensin I to angiotensin II. Angiotensin II raises blood pressure by constricting blood vessels. It triggers the release of aldosterone from the adrenal cortex, which in turn stimulates the renal tubules to reabsorb more sodium. Angiotensin II also triggers the release of anti-diuretic hormone from the hypothalamus, which leads to water retention. It acts directly on the nephrons and decreases GFR.

### **Glossary**

angiotensin converting enzyme (ACE)

enzyme that converts angiotensin I to angiotensin II

angiotensin I

product in the renin-angiotensin-aldosterone pathway

angiotensin II

molecule that affects different organs to increase blood pressure

anti-diuretic hormone (ADH)

hormone that prevents the loss of water

renin-angiotensin-aldosterone

biochemical pathway that activates angiotensin II, which increases blood pressure

vasodilator

compound that increases the diameter of blood vessels

vasopressin

another name for anti-diuretic hormone

## Introduction

class="introduction"

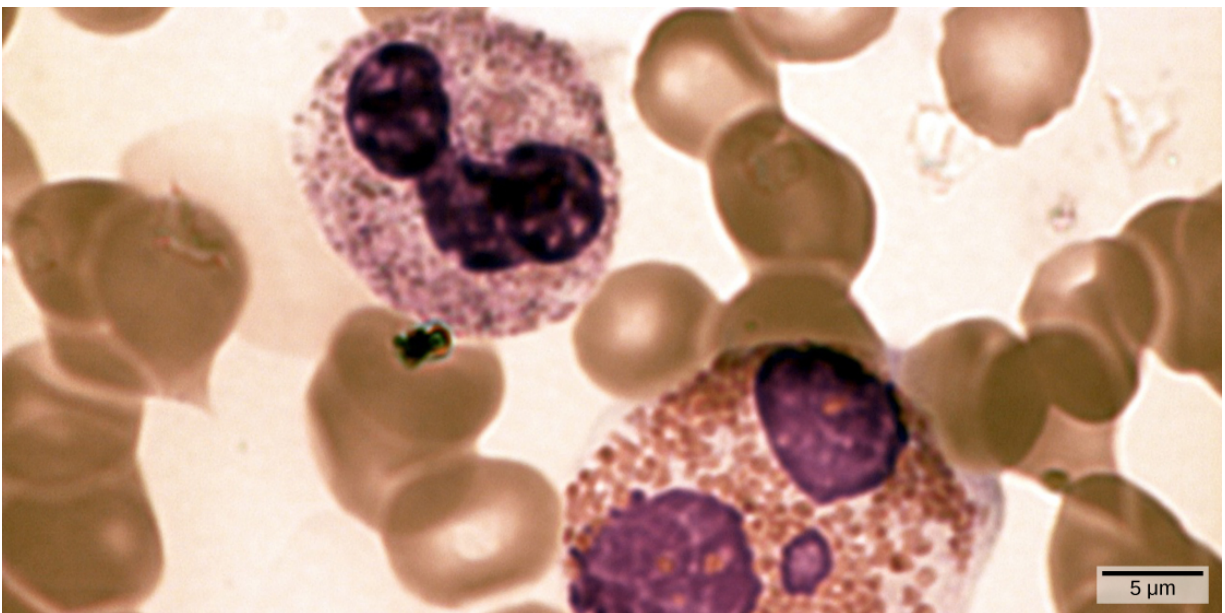
In this  
compound  
light  
micrograph  
purple-  
stained  
neutrophil  
(upper left)  
and  
eosinophil  
(lower right)  
are white  
blood cells  
that float  
among red  
blood cells  
in this blood  
smear.

Neutrophils  
provide an  
early, rapid,  
and  
nonspecific  
defense  
against  
invading  
pathogens.

Eosinophils  
play a  
variety of  
roles in the  
immune  
response.  
Red blood



cells are  
about 7–8  
 $\mu\text{m}$  in  
diameter,  
and a  
neutrophil is  
about 10–  
12 $\mu\text{m}$ .  
(credit:  
modificatio  
n of work  
by Dr.  
David  
Csaba)



The environment consists of numerous **pathogens**, which are agents, usually microorganisms, that cause diseases in their hosts. A **host** is the organism that is invaded and often harmed by a pathogen. Pathogens include bacteria, protists, fungi and other infectious organisms. We are constantly exposed to pathogens in food and water, on surfaces, and in the air. Mammalian immune systems evolved for protection from such pathogens; they are composed of an extremely diverse array of specialized

cells and soluble molecules that coordinate a rapid and flexible defense system capable of providing protection from a majority of these disease agents.

Components of the immune system constantly search the body for signs of pathogens. When pathogens are found, immune factors are mobilized to the site of an infection. The immune factors identify the nature of the pathogen, strengthen the corresponding cells and molecules to combat it efficiently, and then halt the immune response after the infection is cleared to avoid unnecessary host cell damage. The immune system can remember pathogens to which it has been exposed to create a more efficient response upon re-exposure. This memory can last several decades. Features of the immune system, such as pathogen identification, specific response, amplification, retreat, and remembrance are essential for survival against pathogens. The immune response can be classified as either innate or active. The innate immune response is always present and attempts to defend against all pathogens rather than focusing on specific ones. Conversely, the adaptive immune response stores information about past infections and mounts pathogen-specific defenses.

## **Glossary**

pathogen

an agent, usually a microorganism, that causes disease in the organisms that they invade

host

an organism that is invaded by a pathogen or parasite

## Innate Immune Response

By the end of this section, you will be able to:

- Describe physical and chemical immune barriers
- Explain immediate and induced innate immune responses
- Discuss natural killer cells
- Describe major histocompatibility class I molecules
- Summarize how the proteins in a complement system function to destroy extracellular pathogens

The immune system comprises both innate and adaptive immune responses. **Innate immunity** occurs naturally because of genetic factors or physiology; it is not induced by infection or vaccination but works to reduce the workload for the adaptive immune response. Both the innate and adaptive levels of the immune response involve secreted proteins, receptor-mediated signaling, and intricate cell-to-cell communication. The innate immune system developed early in animal evolution, roughly a billion years ago, as an essential response to infection. Innate immunity has a limited number of specific targets: any pathogenic threat triggers a consistent sequence of events that can identify the type of pathogen and either clear the infection independently or mobilize a highly specialized adaptive immune response. For example, tears and mucus secretions contain microbicidal factors.

## Physical and Chemical Barriers

Before any immune factors are triggered, the skin functions as a continuous, impassable barrier to potentially infectious pathogens. Pathogens are killed or inactivated on the skin by desiccation (drying out) and by the skin's acidity. In addition, beneficial microorganisms that coexist on the skin compete with invading pathogens, preventing infection. Regions of the body that are not protected by skin (such as the eyes and mucus membranes) have alternative methods of defense, such as tears and mucus secretions that trap and rinse away pathogens, and cilia in the nasal passages and respiratory tract that push the mucus with the pathogens out of the body. Throughout the body are other defenses, such as the low pH of the stomach (which inhibits the growth of pathogens), blood proteins that bind

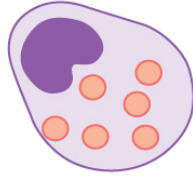
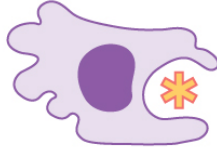
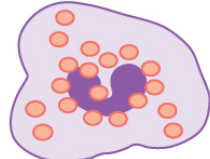
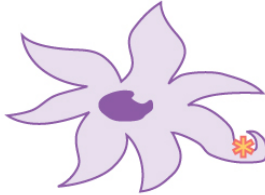


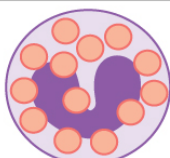
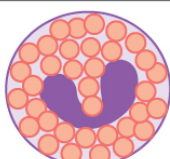
and disrupt bacterial cell membranes, and the process of urination (which flushes pathogens from the urinary tract).

Despite these barriers, pathogens may enter the body through skin abrasions or punctures, or by collecting on mucosal surfaces in large numbers that overcome the mucus or cilia. Some pathogens have evolved specific mechanisms that allow them to overcome physical and chemical barriers. When pathogens do enter the body, the innate immune system responds with inflammation, pathogen engulfment, and secretion of immune factors and proteins.

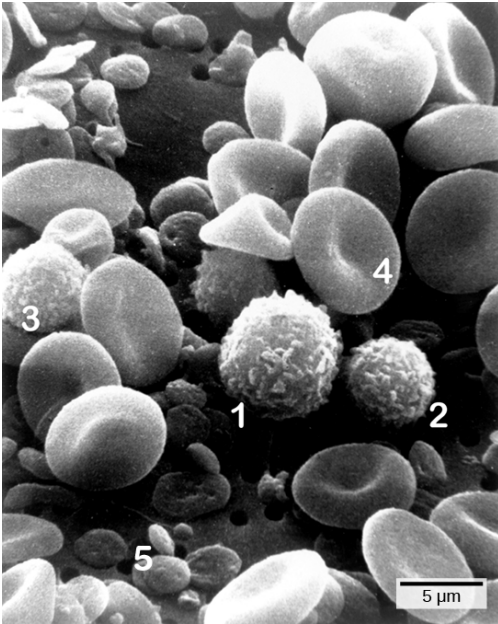
## Pathogen Recognition

An infection may be intracellular or extracellular, depending on the pathogen. All viruses infect cells and replicate within those cells (intracellularly), whereas bacteria and other parasites may replicate intracellularly or extracellularly, depending on the species. The innate immune system must respond accordingly: by identifying the extracellular pathogen and/or by identifying host cells that have already been infected. When a pathogen enters the body, cells in the blood and lymph detect the specific **pathogen-associated molecular patterns (PAMPs)** on the pathogen's surface. PAMPs are carbohydrate, polypeptide, and nucleic acid "signatures" that are expressed by viruses, bacteria, and parasites but which differ from molecules on host cells. The immune system has specific cells, described in [\[link\]](#) and shown in [\[link\]](#), with receptors that recognize these PAMPs. A **macrophage** is a large phagocytic cell that engulfs foreign particles and pathogens. Macrophages recognize PAMPs via complementary **pattern recognition receptors (PRRs)**. PRRs are molecules on macrophages and dendritic cells which are in contact with the external environment. A **monocyte** is a type of white blood cell that circulates in the blood and lymph and differentiates into macrophages after it moves into infected tissue. Dendritic cells bind molecular signatures of pathogens and promote pathogen engulfment and destruction. Toll-like receptors (TLRs) are a type of PRR that recognizes molecules that are shared by pathogens but distinguishable from host molecules). TLRs are present in invertebrates as well as vertebrates, and appear to be one of the

most ancient components of the immune system. TLRs have also been identified in the mammalian nervous system.

Cell type	Characteristics	Location	Image
Mast cell	Dilates blood vessels and induces inflammation through release of histamines and heparin. Recruits macrophages and neutrophils. Involved in wound healing and defense against pathogens but can also be responsible for allergic reactions.	Connective tissues, mucous membranes	
Macrophage	Phagocytic cell that consumes foreign pathogens and cancer cells. Stimulates response of other immune cells.	Migrates from blood vessels into tissues.	
Natural killer cell	Kills tumor cells and virus-infected cells.	Circulates in blood and migrates into tissues.	
Dendritic cell	Presents antigens on its surface, thereby triggering adaptive immunity.	Present in epithelial tissue, including skin, lung and tissues of the digestive tract. Migrates to lymph nodes upon activation.	
Monocyte	Differentiates into macrophages and dendritic cells in response to inflammation.	Stored in spleen, moves through blood vessels to infected tissues.	
Neutrophil	First responders at the site of infection or trauma, this abundant phagocytic cell represents 50-60 percent of all leukocytes. Releases toxins that kill or inhibit bacteria and fungi and recruits other immune cells to the site of infection.	Migrates from blood vessels into tissues.	
Basophil	Responsible for defense against parasites. Releases histamines that cause inflammation and may be responsible for allergic reactions.	Circulates in blood and migrates to tissues.	
Eosinophil	Releases toxins that kill bacteria and parasites but also causes tissue damage.	Circulates in blood and migrates to tissues.	

The characteristics and location of cells involved in the innate immune system are described. (credit: modification of work by NIH)



Cells of the blood include  
(1) monocytes, (2)  
lymphocytes, (3)  
neutrophils, (4) red blood  
cells, and (5) platelets.

Note the very similar  
morphologies of the  
leukocytes (1, 2, 3).  
(credit: modification of  
work by Bruce Wetzel,  
Harry Schaefer, NCI;  
scale-bar data from Matt  
Russell)

## Cytokine Release Effect

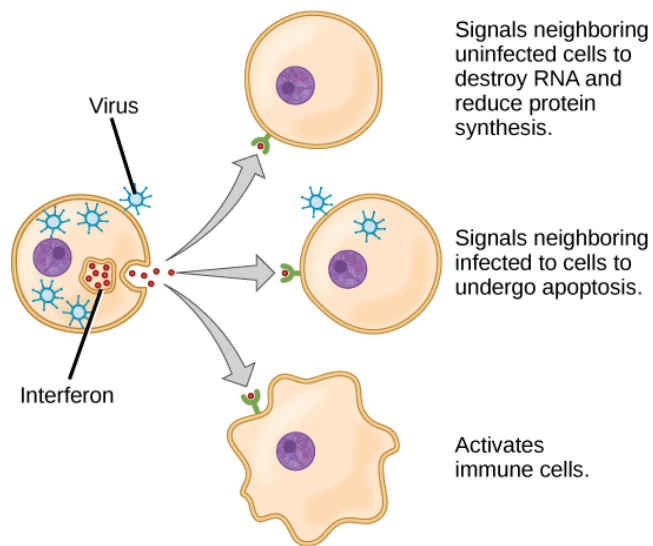
The binding of PRRs with PAMPs triggers the release of cytokines, which signal that a pathogen is present and needs to be destroyed along with any infected cells. A **cytokine** is a chemical messenger that regulates cell differentiation (form and function), proliferation (production), and gene expression to affect immune responses. At least 40 types of cytokines exist in humans that differ in terms of the cell type that produces them, the cell type that responds to them, and the changes they produce. One type cytokine, interferon, is illustrated in [\[link\]](#).

One subclass of cytokines is the interleukin (IL), so named because they mediate interactions between leukocytes (white blood cells). Interleukins are involved in bridging the innate and adaptive immune responses. In addition to being released from cells after PAMP recognition, cytokines are released by the infected cells which bind to nearby uninfected cells and induce those cells to release cytokines, which results in a cytokine burst.

A second class of early-acting cytokines is interferons, which are released by infected cells as a warning to nearby uninfected cells. One of the functions of an **interferon** is to inhibit viral replication. They also have other important functions, such as tumor surveillance. Interferons work by signaling neighboring uninfected cells to destroy RNA and reduce protein synthesis, signaling neighboring infected cells to undergo apoptosis (programmed cell death), and activating immune cells.

In response to interferons, uninfected cells alter their gene expression, which increases the cells' resistance to infection. One effect of interferon-induced gene expression is a sharply reduced cellular protein synthesis. Virally infected cells produce more viruses by synthesizing large quantities of viral proteins. Thus, by reducing protein synthesis, a cell becomes resistant to viral infection.



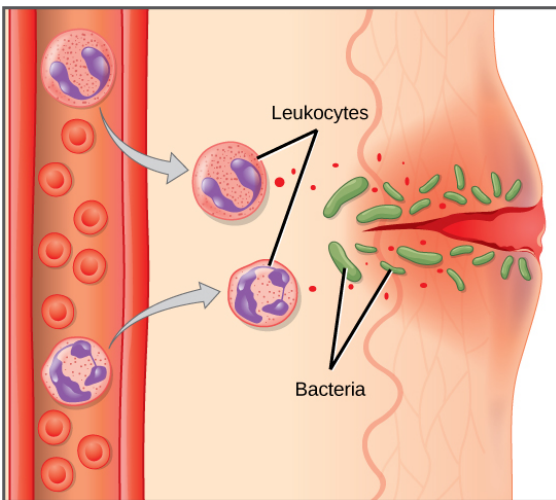


Interferons are cytokines that are released by a cell infected with a virus. Response of neighboring cells to interferon helps stem the infection.

## Phagocytosis and Inflammation

The first cytokines to be produced are pro-inflammatory; that is, they encourage **inflammation**, the localized redness, swelling, heat, and pain that result from the movement of leukocytes and fluid through increasingly permeable capillaries to a site of infection. The population of leukocytes that arrives at an infection site depends on the nature of the infecting pathogen. Both macrophages and dendritic cells engulf pathogens and cellular debris through phagocytosis. A **neutrophil** is also a phagocytic leukocyte that engulfs and digests pathogens. Neutrophils, shown in [\[link\]](#), are the most abundant leukocytes of the immune system. Neutrophils have a nucleus with two to five lobes, and they contain organelles, called lysosomes, that digest engulfed pathogens. An **eosinophil** is a leukocyte that works with other eosinophils to surround a parasite; it is involved in the allergic response and in protection against helminthes (parasitic worms).

Neutrophils and eosinophils are particularly important leukocytes that engulf large pathogens, such as bacteria and fungi. A **mast cell** is a leukocyte that produces inflammatory molecules, such as histamine, in response to large pathogens. A **basophil** is a leukocyte that, like a neutrophil, releases chemicals to stimulate the inflammatory response as illustrated in [\[link\]](#). Basophils are also involved in allergy and hypersensitivity responses and induce specific types of inflammatory responses. Eosinophils and basophils produce additional inflammatory mediators to recruit more leukocytes. A hypersensitive immune response to harmless antigens, such as in pollen, often involves the release of histamine by basophils and mast cells.



In response to a cut, mast cells secrete histamines that cause nearby capillaries to dilate. Neutrophils and monocytes leave the capillaries. Monocytes mature into macrophages. Neutrophils, dendritic cells and macrophages release chemicals to stimulate the inflammatory response. Neutrophils and macrophages

also consume invading bacteria by phagocytosis.

Cytokines also send feedback to cells of the nervous system to bring about the overall symptoms of feeling sick, which include lethargy, muscle pain, and nausea. These effects may have evolved because the symptoms encourage the individual to rest and prevent them from spreading the infection to others. Cytokines also increase the core body temperature, causing a fever, which causes the liver to withhold iron from the blood. Without iron, certain pathogens, such as some bacteria, are unable to replicate; this is called nutritional immunity.

**Note:**

Link to Learning

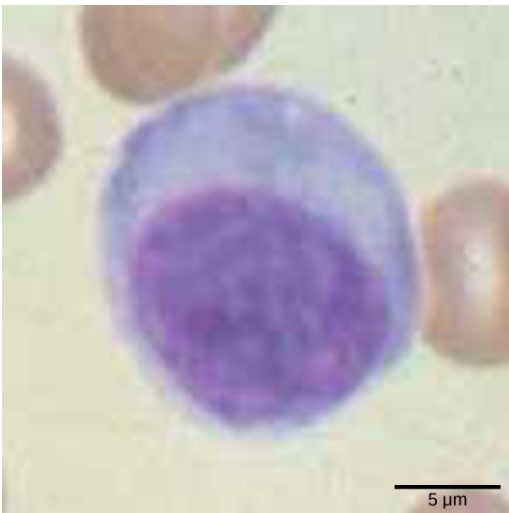


Watch this 23-second stop-motion [video](#) showing a neutrophil that searches for and engulfs fungus spores during an elapsed time of about 79 minutes.

## Natural Killer Cells

**Lymphocytes** are leukocytes that are histologically identifiable by their large, darkly staining nuclei; they are small cells with very little cytoplasm, as shown in [\[link\]](#). Infected cells are identified and destroyed by **natural killer (NK) cells**, lymphocytes that can kill cells infected with viruses or tumor cells (abnormal cells that uncontrollably divide and invade other tissue). T cells and B cells of the adaptive immune system also are

classified as lymphocytes. **T cells** are lymphocytes that mature in the thymus gland, and **B cells** are lymphocytes that mature in the bone marrow. NK cells identify intracellular infections, especially from viruses, by the altered expression of **major histocompatibility class (MHC) I molecules** on the surface of infected cells. MHC I molecules are proteins on the surfaces of all nucleated cells, thus they are scarce on red blood cells and platelets which are non-nucleated. The function of MHC I molecules is to display fragments of proteins from the infectious agents within the cell to T-cells; healthy cells will be ignored, while “non-self” or foreign proteins will be attacked by the immune system. MHC II molecules are found mainly on cells containing antigens (“non-self proteins”) and on lymphocytes. **MHC II molecules** interact with helper T-cells to trigger the appropriate immune response, which may include the inflammatory response.



Lymphocytes, such as NK cells, are characterized by their large nuclei that actively absorb Wright stain and therefore appear dark colored under a microscope.

An infected cell (or a tumor cell) is usually incapable of synthesizing and displaying MHC I molecules appropriately. The metabolic resources of cells infected by some viruses produce proteins that interfere with MHC I processing and/or trafficking to the cell surface. The reduced MHC I on host cells varies from virus to virus and results from active inhibitors being produced by the viruses. This process can deplete host MHC I molecules on the cell surface, which NK cells detect as “unhealthy” or “abnormal” while searching for cellular MHC I molecules. Similarly, the dramatically altered gene expression of tumor cells leads to expression of extremely deformed or absent MHC I molecules that also signal “unhealthy” or “abnormal.”

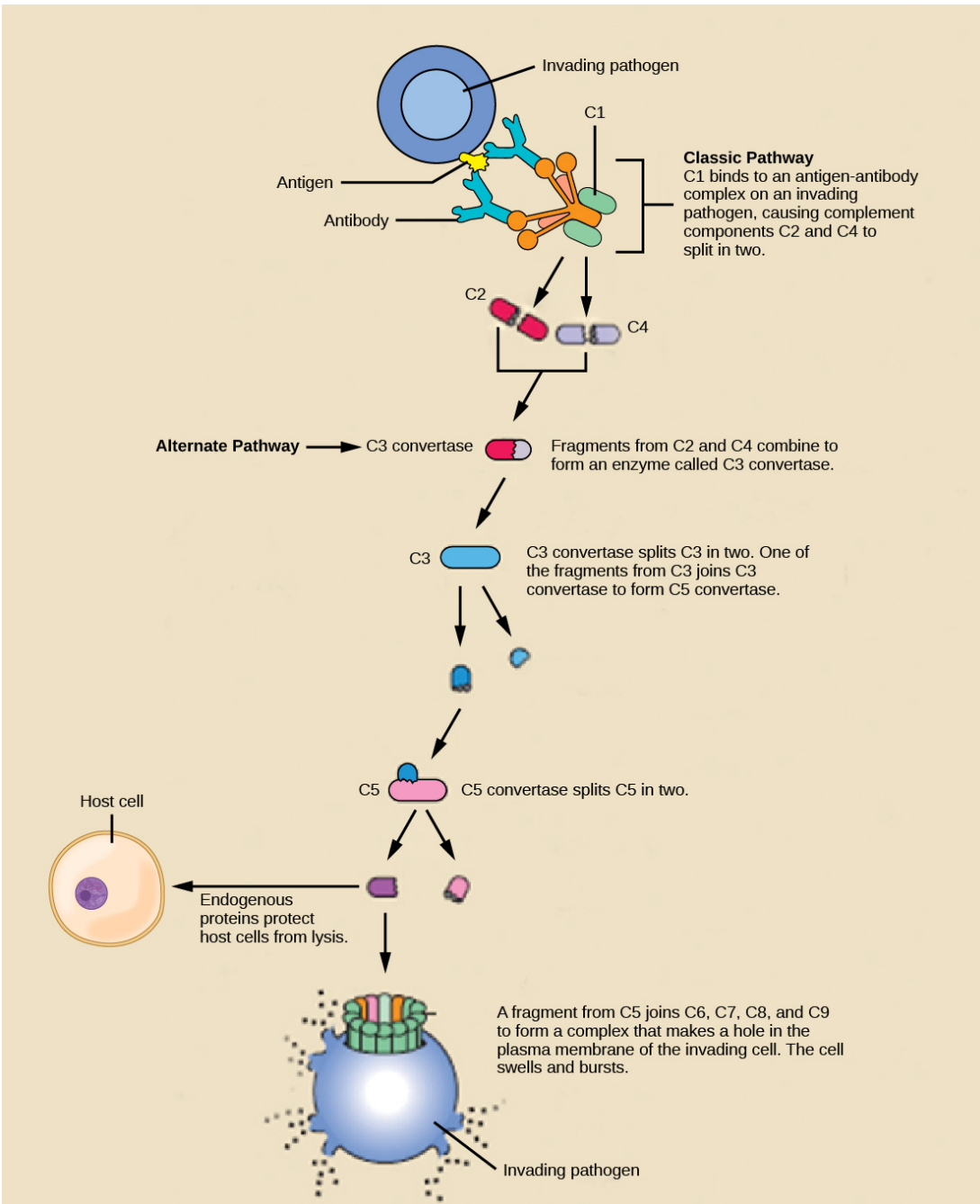
NK cells are always active; an interaction with normal, intact MHC I molecules on a healthy cell disables the killing sequence, and the NK cell moves on. After the NK cell detects an infected or tumor cell, its cytoplasm secretes granules comprised of **perforin**, a destructive protein that creates a pore in the target cell. Granzymes are released along with the perforin in the immunological synapse. A **granzyme** is a protease that digests cellular proteins and induces the target cell to undergo programmed cell death, or apoptosis. Phagocytic cells then digest the cell debris left behind. NK cells are constantly patrolling the body and are an effective mechanism for controlling potential infections and preventing cancer progression.

## Complement

An array of approximately 20 types of soluble proteins, called a **complement system**, functions to destroy extracellular pathogens. Cells of the liver and macrophages synthesize complement proteins continuously; these proteins are abundant in the blood serum and are capable of responding immediately to infecting microorganisms. The complement system is so named because it is complementary to the antibody response of the adaptive immune system. Complement proteins bind to the surfaces of microorganisms and are particularly attracted to pathogens that are already bound by antibodies. Binding of complement proteins occurs in a specific and highly regulated sequence, with each successive protein being activated by cleavage and/or structural changes induced upon binding of the preceding protein(s). After the first few complement proteins bind, a

cascade of sequential binding events follows in which the pathogen rapidly becomes coated in complement proteins.

Complement proteins perform several functions. The proteins serve as a marker to indicate the presence of a pathogen to phagocytic cells, such as macrophages and B cells, and enhance engulfment; this process is called **opsonization**. Certain complement proteins can combine to form attack complexes that open pores in microbial cell membranes. These structures destroy pathogens by causing their contents to leak, as illustrated in [\[link\]](#).



The classic pathway for the complement cascade involves the attachment of several initial complement proteins to an antibody-bound pathogen followed by rapid activation and binding of many more complement proteins and the creation of destructive pores in the microbial cell envelope and cell wall. The alternate pathway does not involve antibody activation. Rather, C3 convertase spontaneously breaks down C3.

Endogenous regulatory proteins prevent the complement complex from binding to host cells. Pathogens lacking these regulatory proteins are lysed. (credit: modification of work by NIH)

## Section Summary

The innate immune system serves as a first responder to pathogenic threats that bypass natural physical and chemical barriers of the body. Using a combination of cellular and molecular attacks, the innate immune system identifies the nature of a pathogen and responds with inflammation, phagocytosis, cytokine release, destruction by NK cells, and/or a complement system. When innate mechanisms are insufficient to clear an infection, the adaptive immune response is informed and mobilized.

## Review Questions

### Exercise:

#### Problem:

Which of the following is a barrier against pathogens provided by the skin?

- a. high pH
- b. mucus
- c. tears
- d. desiccation

---

#### Solution:

D

### Exercise:



**Problem:**

Although interferons have several effects, they are particularly useful against infections with which type of pathogen?

- a. bacteria
- b. viruses
- c. fungi
- d. helminths

---

**Solution:**

B

**Exercise:****Problem:**

Which organelle do phagocytes use to digest engulfed particles?

- a. lysosome
- b. nucleus
- c. endoplasmic reticulum
- d. mitochondria

---

**Solution:**

A

**Exercise:****Problem:**

Which innate immune system component uses MHC I molecules directly in its defense strategy?

- a. macrophages
- b. neutrophils

- c. NK cells
  - d. interferon
- 

**Solution:**

C

## Free Response

**Exercise:**

**Problem:**

Different MHC I molecules between donor and recipient cells can lead to rejection of a transplanted organ or tissue. Suggest a reason for this.

---

**Solution:**

If the MHC I molecules expressed on donor cells differ from the MHC I molecules expressed on recipient cells, NK cells may identify the donor cells as “non-self” and produce perforin and granzymes to induce the donor cells to undergo apoptosis, which would destroy the transplanted organ.

**Exercise:**

**Problem:**

If a series of genetic mutations prevented some, but not all, of the complement proteins from binding antibodies or pathogens, would the entire complement system be compromised?

---

**Solution:**

The entire complement system would probably be affected even when only a few members were mutated such that they could no longer bind. Because the complement involves the binding of activated proteins in a specific sequence, when one or more proteins in the sequence are

absent, the subsequent proteins would be incapable of binding to elicit the complement's pathogen-destructive effects.

## **Glossary**

### **basophil**

leukocyte that releases chemicals usually involved in the inflammatory response

### **B cell**

lymphocyte that matures in the bone marrow and differentiates into antibody-secreting plasma cells

### **complement system**

array of approximately 20 soluble proteins of the innate immune system that enhance phagocytosis, bore holes in pathogens, and recruit lymphocytes; enhances the adaptive response when antibodies are produced

### **cytokine**

chemical messenger that regulates cell differentiation, proliferation, gene expression, and cell trafficking to effect immune responses

### **eosinophil**

leukocyte that responds to parasites and is involved in the allergic response

### **granzyme**

protease that enters target cells through perforin and induces apoptosis in the target cells; used by NK cells and killer T cells

### **inflammation**

localized redness, swelling, heat, and pain that results from the movement of leukocytes and fluid through opened capillaries to a site of infection

### **innate immunity**

immunity that occurs naturally because of genetic factors or physiology, and is not induced by infection or vaccination

interferon

cytokine that inhibits viral replication and modulates the immune response

lymphocyte

leukocyte that is histologically identifiable by its large nuclei; it is a small cell with very little cytoplasm

macrophage

large phagocytic cell that engulfs foreign particles and pathogens

major histocompatibility class (MHC) I/II molecule

protein found on the surface of all nucleated cells (I) or specifically on antigen-presenting cells (II) that signals to immune cells whether the cell is healthy/normal or is infected/cancerous; it provides the appropriate template into which antigens can be loaded for recognition by lymphocytes

mast cell

leukocyte that produces inflammatory molecules, such as histamine, in response to large pathogens and allergens

monocyte

type of white blood cell that circulates in the blood and lymph and differentiates into macrophages after it moves into infected tissue

natural killer (NK) cell

lymphocyte that can kill cells infected with viruses or tumor cells

neutrophil

phagocytic leukocyte that engulfs and digests pathogens

opsonization

process that enhances phagocytosis using proteins to indicate the presence of a pathogen to phagocytic cells

pathogen-associated molecular pattern (PAMP)

carbohydrate, polypeptide, and nucleic acid “signature” that is expressed by viruses, bacteria, and parasites but differs from molecules on host cells

pattern recognition receptor (PRR)

molecule on macrophages and dendritic cells that binds molecular signatures of pathogens and promotes pathogen engulfment and destruction

perforin

destructive protein that creates a pore in the target cell; used by NK cells and killer T cells

T cell

lymphocyte that matures in the thymus gland; one of the main cells involved in the adaptive immune system

## Adaptive Immune Response

By the end of this section, you will be able to:

- Explain adaptive immunity
- Compare and contrast adaptive and innate immunity
- Describe cell-mediated immune response and humoral immune response
- Describe immune tolerance

The adaptive, or acquired, immune response takes days or even weeks to become established—much longer than the innate response; however, adaptive immunity is more specific to pathogens and has memory.

**Adaptive immunity** is an immunity that occurs after exposure to an antigen either from a pathogen or a vaccination. This part of the immune system is activated when the innate immune response is insufficient to control an infection. In fact, without information from the innate immune system, the adaptive response could not be mobilized. There are two types of adaptive responses: the **cell-mediated immune response**, which is carried out by T cells, and the **humoral immune response**, which is controlled by activated B cells and antibodies. Activated T cells and B cells that are specific to molecular structures on the pathogen proliferate and attack the invading pathogen. Their attack can kill pathogens directly or secrete antibodies that enhance the phagocytosis of pathogens and disrupt the infection. Adaptive immunity also involves a memory to provide the host with long-term protection from reinfection with the same type of pathogen; on re-exposure, this memory will facilitate an efficient and quick response.

## Antigen-presenting Cells

Unlike NK cells of the innate immune system, B cells (B lymphocytes) are a type of white blood cell that gives rise to antibodies, whereas T cells (T lymphocytes) are a type of white blood cell that plays an important role in the immune response. T cells are a key component in the cell-mediated response—the specific immune response that utilizes T cells to neutralize cells that have been infected with viruses and certain bacteria. There are three types of T cells: cytotoxic, helper, and suppressor T cells. Cytotoxic T cells destroy virus-infected cells in the cell-mediated immune response, and

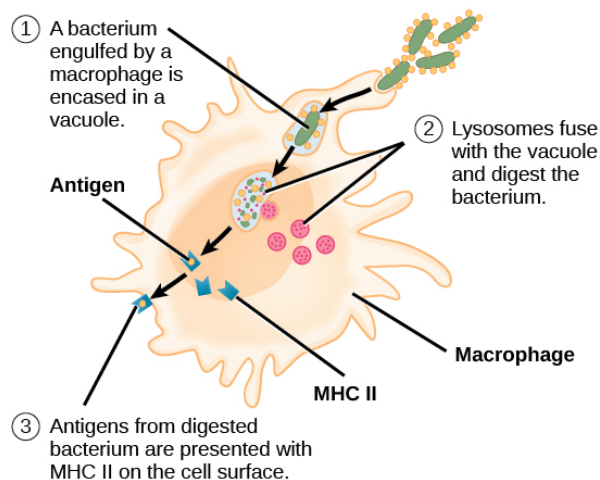
helper T cells play a part in activating both the antibody and the cell-mediated immune responses. Suppressor T cells deactivate T cells and B cells when needed, and thus prevent the immune response from becoming too intense.

An **antigen** is a foreign or “non-self” macromolecule that reacts with cells of the immune system. Not all antigens will provoke a response. For instance, individuals produce innumerable “self” antigens and are constantly exposed to harmless foreign antigens, such as food proteins, pollen, or dust components. The suppression of immune responses to harmless macromolecules is highly regulated and typically prevents processes that could be damaging to the host, known as tolerance.

The innate immune system contains cells that detect potentially harmful antigens, and then inform the adaptive immune response about the presence of these antigens. An **antigen-presenting cell (APC)** is an immune cell that detects, engulfs, and informs the adaptive immune response about an infection. When a pathogen is detected, these APCs will phagocytose the pathogen and digest it to form many different fragments of the antigen. Antigen fragments will then be transported to the surface of the APC, where they will serve as an indicator to other immune cells. **Dendritic cells** are immune cells that process antigen material; they are present in the skin (Langerhans cells) and the lining of the nose, lungs, stomach, and intestines. Sometimes a dendritic cell presents on the surface of other cells to induce an immune response, thus functioning as an antigen-presenting cell. Macrophages also function as APCs. Before activation and differentiation, B cells can also function as APCs.

After phagocytosis by APCs, the phagocytic vesicle fuses with an intracellular lysosome forming phagolysosome. Within the phagolysosome, the components are broken down into fragments; the fragments are then loaded onto MHC class I or MHC class II molecules and are transported to the cell surface for antigen presentation, as illustrated in [\[link\]](#). Note that T lymphocytes cannot properly respond to the antigen unless it is processed and embedded in an MHC II molecule. APCs express MHC on their surfaces, and when combined with a foreign antigen, these complexes signal a “non-self” invader. Once the fragment of antigen is embedded in

the MHC II molecule, the immune cell can respond. Helper T- cells are one of the main lymphocytes that respond to antigen-presenting cells. Recall that all other nucleated cells of the body expressed MHC I molecules, which signal “healthy” or “normal.”



An APC, such as a macrophage, engulfs and digests a foreign bacterium. An antigen from the bacterium is presented on the cell surface in conjunction with an MHC II molecule. Lymphocytes of the adaptive immune response interact with antigen-embedded MHC II molecules to mature into functional immune cells.

**Note:**

Link to Learning



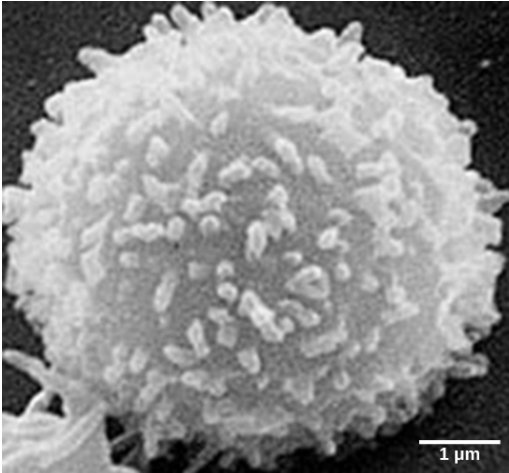


This [animation](#) from Rockefeller University shows how dendritic cells act as sentinels in the body's immune system.

## T and B Lymphocytes

Lymphocytes in human circulating blood are approximately 80 to 90 percent T cells, shown in [\[link\]](#), and 10 to 20 percent B cells. Recall that the T cells are involved in the cell-mediated immune response, whereas B cells are part of the humoral immune response.

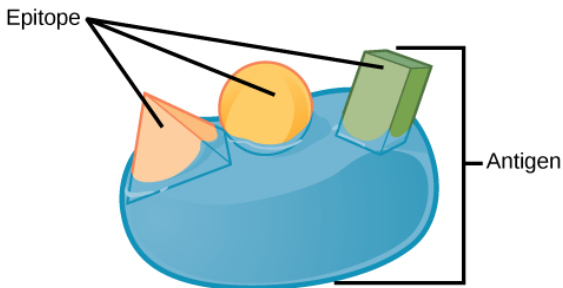
T cells encompass a heterogeneous population of cells with extremely diverse functions. Some T cells respond to APCs of the innate immune system, and indirectly induce immune responses by releasing cytokines. Other T cells stimulate B cells to prepare their own response. Another population of T cells detects APC signals and directly kills the infected cells. Other T cells are involved in suppressing inappropriate immune reactions to harmless or “self” antigens.



This scanning electron micrograph shows a T lymphocyte, which is responsible for the cell-mediated immune response. T cells are able to recognize antigens. (credit: modification of work by NCI; scale-bar data from Matt Russell)

T and B cells exhibit a common theme of recognition/binding of specific antigens via a complementary receptor, followed by activation and self-amplification/maturation to specifically bind to the particular antigen of the infecting pathogen. T and B lymphocytes are also similar in that each cell only expresses one type of antigen receptor. Any individual may possess a population of T and B cells that together express a near limitless variety of antigen receptors that are capable of recognizing virtually any infecting pathogen. T and B cells are activated when they recognize small components of antigens, called **epitopes**, presented by APCs, illustrated in [\[link\]](#). Note that recognition occurs at a specific epitope rather than on the entire antigen; for this reason, epitopes are known as “antigenic determinants.” In the absence of information from APCs, T and B cells remain inactive, or naïve, and are unable to prepare an immune response.

The requirement for information from the APCs of innate immunity to trigger B cell or T cell activation illustrates the essential nature of the innate immune response to the functioning of the entire immune system.



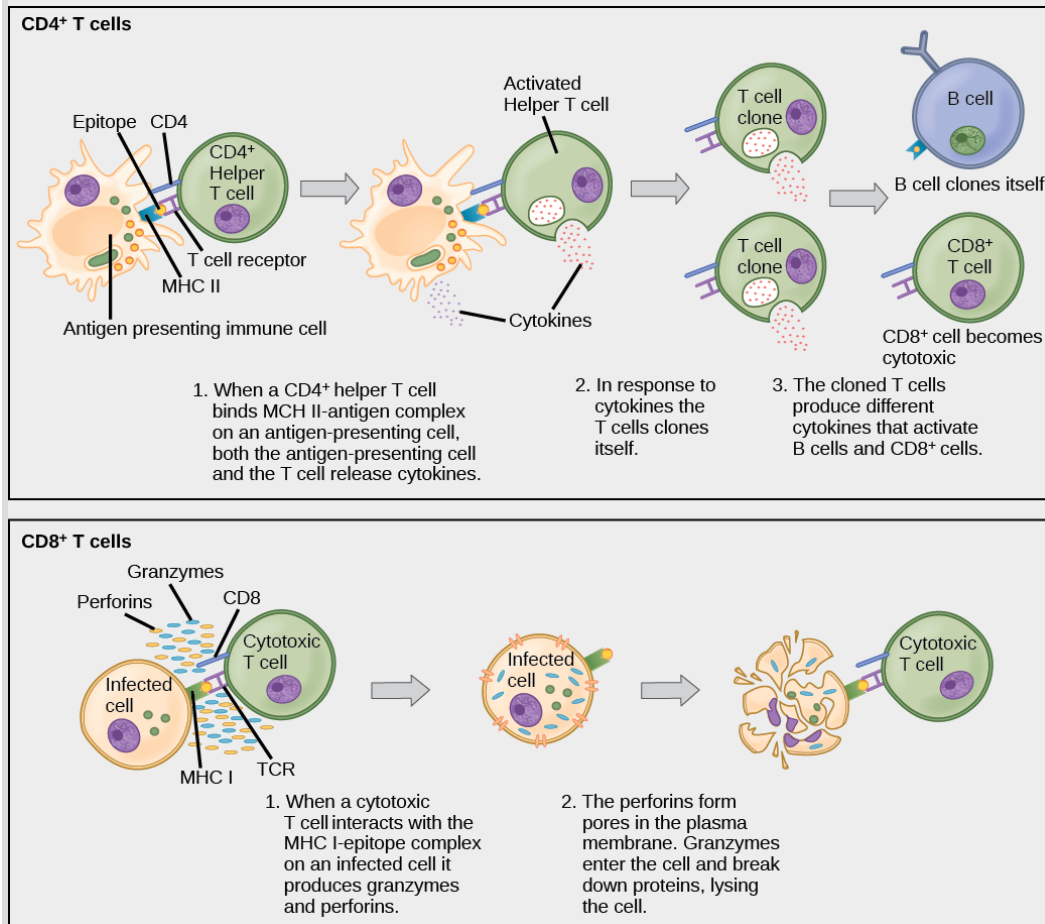
An antigen is a macromolecule that reacts with components of the immune system. A given antigen may contain several motifs that are recognized by immune cells. Each motif is an epitope. In this figure, the entire structure is an antigen, and the orange, salmon and green components projecting from it represent potential epitopes.

Naïve T cells can express one of two different molecules, CD4 or CD8, on their surface, as shown in [\[link\]](#), and are accordingly classified as CD4<sup>+</sup> or CD8<sup>+</sup> cells. These molecules are important because they regulate how a T cell will interact with and respond to an APC. Naïve CD4<sup>+</sup> cells bind APCs via their antigen-embedded MHC II molecules and are stimulated to become **helper T (T<sub>H</sub>) lymphocytes**, cells that go on to stimulate B cells (or cytotoxic T cells) directly or secrete cytokines to inform more and various target cells about the pathogenic threat. In contrast, CD8<sup>+</sup> cells

engage antigen-embedded MHC I molecules on APCs and are stimulated to become **cytotoxic T lymphocytes (CTLs)**, which directly kill infected cells by apoptosis and emit cytokines to amplify the immune response. The two populations of T cells have different mechanisms of immune protection, but both bind MHC molecules via their antigen receptors called T cell receptors (TCRs). The CD4 or CD8 surface molecules differentiate whether the TCR will engage an MHC II or an MHC I molecule. Because they assist in binding specificity, the CD4 and CD8 molecules are described as coreceptors.

## Note:

### Art Connection



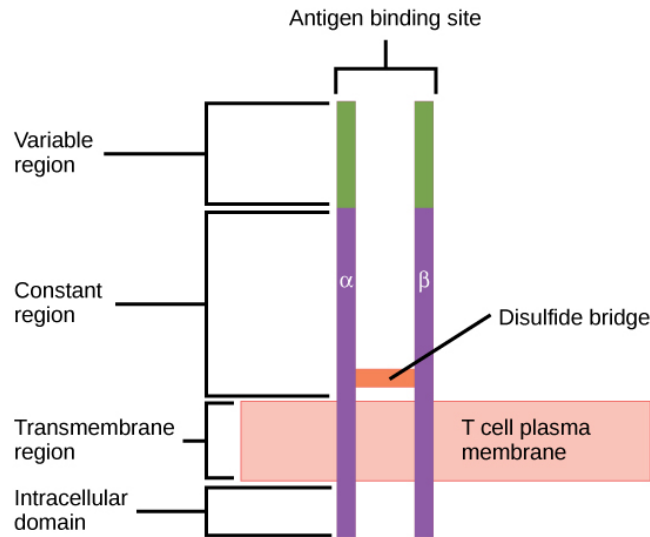
Naïve CD4<sup>+</sup> T cells engage MHC II molecules on antigen-presenting cells (APCs) and become activated. Clones of the

activated helper T cell, in turn, activate B cells and CD8<sup>+</sup> T cells, which become cytotoxic T cells. Cytotoxic T cells kill infected cells.

Which of the following statements about T cells is false?

- a. Helper T cells release cytokines while cytotoxic T cells kill the infected cell.
- b. Helper T cells are CD4<sup>+</sup>, while cytotoxic T cells are CD8<sup>+</sup>.
- c. MHC II is a receptor found on most body cells, while MHC I is a receptor found on immune cells only.
- d. The T cell receptor is found on both CD4<sup>+</sup> and CD8<sup>+</sup> T cells.

Consider the innumerable possible antigens that an individual will be exposed to during a lifetime. The mammalian adaptive immune system is adept in responding appropriately to each antigen. Mammals have an enormous diversity of T cell populations, resulting from the diversity of TCRs. Each TCR consists of two polypeptide chains that span the T cell membrane, as illustrated in [\[link\]](#); the chains are linked by a disulfide bridge. Each polypeptide chain is comprised of a constant domain and a variable domain: a domain, in this sense, is a specific region of a protein that may be regulatory or structural. The intracellular domain is involved in intracellular signaling. A single T cell will express thousands of identical copies of one specific TCR variant on its cell surface. The specificity of the adaptive immune system occurs because it synthesizes millions of different T cell populations, each expressing a TCR that differs in its variable domain. This TCR diversity is achieved by the mutation and recombination of genes that encode these receptors in stem cell precursors of T cells. The binding between an antigen-displaying MHC molecule and a complementary TCR “match” indicates that the adaptive immune system needs to activate and produce that specific T cell because its structure is appropriate to recognize and destroy the invading pathogen.



A T cell receptor spans the membrane and projects variable binding regions into the extracellular space to bind processed antigens via MHC molecules on APCs.

## Helper T Lymphocytes

The  $T_H$  lymphocytes function indirectly to identify potential pathogens for other cells of the immune system. These cells are important for extracellular infections, such as those caused by certain bacteria, helminths, and protozoa.  $T_H$  lymphocytes recognize specific antigens displayed in the MHC II complexes of APCs. There are two major populations of  $T_H$  cells:  $T_H1$  and  $T_H2$ .  $T_H1$  cells secrete cytokines to enhance the activities of macrophages and other T cells.  $T_H1$  cells activate the action of cytotoxic T cells, as well as macrophages.  $T_H2$  cells stimulate naïve B cells to destroy foreign invaders via antibody secretion. Whether a  $T_H1$  or a  $T_H2$  immune response develops depends on the specific types of cytokines secreted by

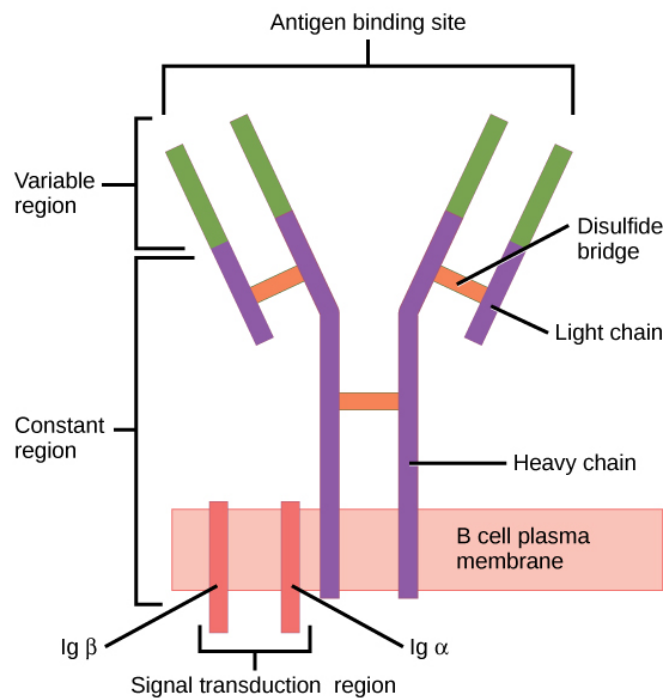
cells of the innate immune system, which in turn depends on the nature of the invading pathogen.

The  $T_H1$ -mediated response involves macrophages and is associated with inflammation. Recall the frontline defenses of macrophages involved in the innate immune response. Some intracellular bacteria, such as *Mycobacterium tuberculosis*, have evolved to multiply in macrophages after they have been engulfed. These pathogens evade attempts by macrophages to destroy and digest the pathogen. When *M. tuberculosis* infection occurs, macrophages can stimulate naïve T cells to become  $T_H1$  cells. These stimulated T cells secrete specific cytokines that send feedback to the macrophage to stimulate its digestive capabilities and allow it to destroy the colonizing *M. tuberculosis*. In the same manner,  $T_H1$ -activated macrophages also become better suited to ingest and kill tumor cells. In summary;  $T_H1$  responses are directed toward intracellular invaders while  $T_H2$  responses are aimed at those that are extracellular.

## B Lymphocytes

When stimulated by the  $T_H2$  pathway, naïve B cells differentiate into antibody-secreting plasma cells. A **plasma cell** is an immune cell that secretes antibodies; these cells arise from B cells that were stimulated by antigens. Similar to T cells, naïve B cells initially are coated in thousands of B cell receptors (BCRs), which are membrane-bound forms of Ig (immunoglobulin, or an antibody). The B cell receptor has two heavy chains and two light chains connected by disulfide linkages. Each chain has a constant and a variable region; the latter is involved in antigen binding. Two other membrane proteins, Ig alpha and Ig beta, are involved in signaling. The receptors of any particular B cell, as shown in [\[link\]](#) are all the same, but the hundreds of millions of different B cells in an individual have distinct recognition domains that contribute to extensive diversity in the types of molecular structures to which they can bind. In this state, B cells function as APCs. They bind and engulf foreign antigens via their BCRs and then display processed antigens in the context of MHC II molecules to  $T_H2$  cells. When a  $T_H2$  cell detects that a B cell is bound to a relevant antigen, it secretes specific cytokines that induce the B cell to

proliferate rapidly, which makes thousands of identical (clonal) copies of it, and then it synthesizes and secretes antibodies with the same antigen recognition pattern as the BCRs. The activation of B cells corresponding to one specific BCR variant and the dramatic proliferation of that variant is known as **clonal selection**. This phenomenon drastically, but briefly, changes the proportions of BCR variants expressed by the immune system, and shifts the balance toward BCRs specific to the infecting pathogen.



B cell receptors are embedded in the membranes of B cells and bind a variety of antigens through their variable regions. The signal transduction region transfers the signal into the cell.

T and B cells differ in one fundamental way: whereas T cells bind antigens that have been digested and embedded in MHC molecules by APCs, B cells



function as APCs that bind intact antigens that have not been processed. Although T and B cells both react with molecules that are termed “antigens,” these lymphocytes actually respond to very different types of molecules. B cells must be able to bind intact antigens because they secrete antibodies that must recognize the pathogen directly, rather than digested remnants of the pathogen. Bacterial carbohydrate and lipid molecules can activate B cells independently from the T cells.

## **Cytotoxic T Lymphocytes**

CTLs, a subclass of T cells, function to clear infections directly. The cell-mediated part of the adaptive immune system consists of CTLs that attack and destroy infected cells. CTLs are particularly important in protecting against viral infections; this is because viruses replicate within cells where they are shielded from extracellular contact with circulating antibodies. When APCs phagocytize pathogens and present MHC I-embedded antigens to naïve  $CD8^+$  T cells that express complementary TCRs, the  $CD8^+$  T cells become activated to proliferate according to clonal selection. These resulting CTLs then identify non-APCs displaying the same MHC I-embedded antigens (for example, viral proteins)—for example, the CTLs identify infected host cells.

Intracellularly, infected cells typically die after the infecting pathogen replicates to a sufficient concentration and lyses the cell, as many viruses do. CTLs attempt to identify and destroy infected cells before the pathogen can replicate and escape, thereby halting the progression of intracellular infections. CTLs also support NK lymphocytes to destroy early cancers. Cytokines secreted by the  $T_H1$  response that stimulates macrophages also stimulate CTLs and enhance their ability to identify and destroy infected cells and tumors.

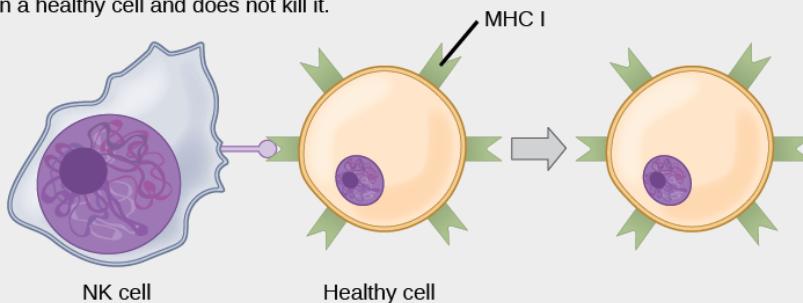
CTLs sense MHC I-embedded antigens by directly interacting with infected cells via their TCRs. Binding of TCRs with antigens activates CTLs to release perforin and granzyme, degradative enzymes that will induce apoptosis of the infected cell. Recall that this is a similar destruction mechanism to that used by NK cells. In this process, the CTL does not

become infected and is not harmed by the secretion of perforin and granzymes. In fact, the functions of NK cells and CTLs are complementary and maximize the removal of infected cells, as illustrated in [\[link\]](#). If the NK cell cannot identify the “missing self” pattern of down-regulated MHC I molecules, then the CTL can identify it by the complex of MHC I with foreign antigens, which signals “altered self.” Similarly, if the CTL cannot detect antigen-embedded MHC I because the receptors are depleted from the cell surface, NK cells will destroy the cell instead. CTLs also emit cytokines, such as interferons, that alter surface protein expression in other infected cells, such that the infected cells can be easily identified and destroyed. Moreover, these interferons can also prevent virally infected cells from releasing virus particles.

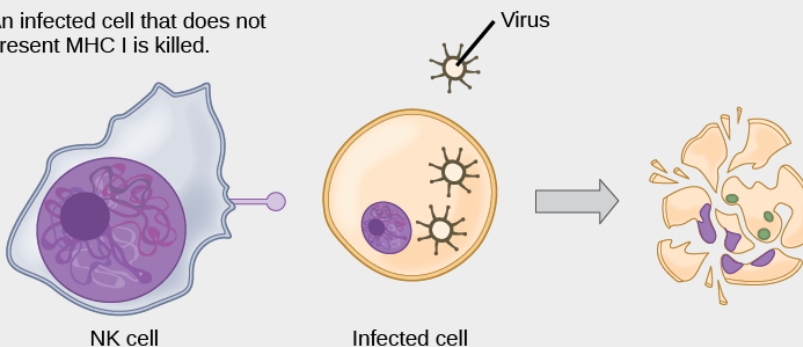
### Note:

#### Art Connection

A natural killer (NK) cell recognizes MHC I on a healthy cell and does not kill it.



An infected cell that does not present MHC I is killed.



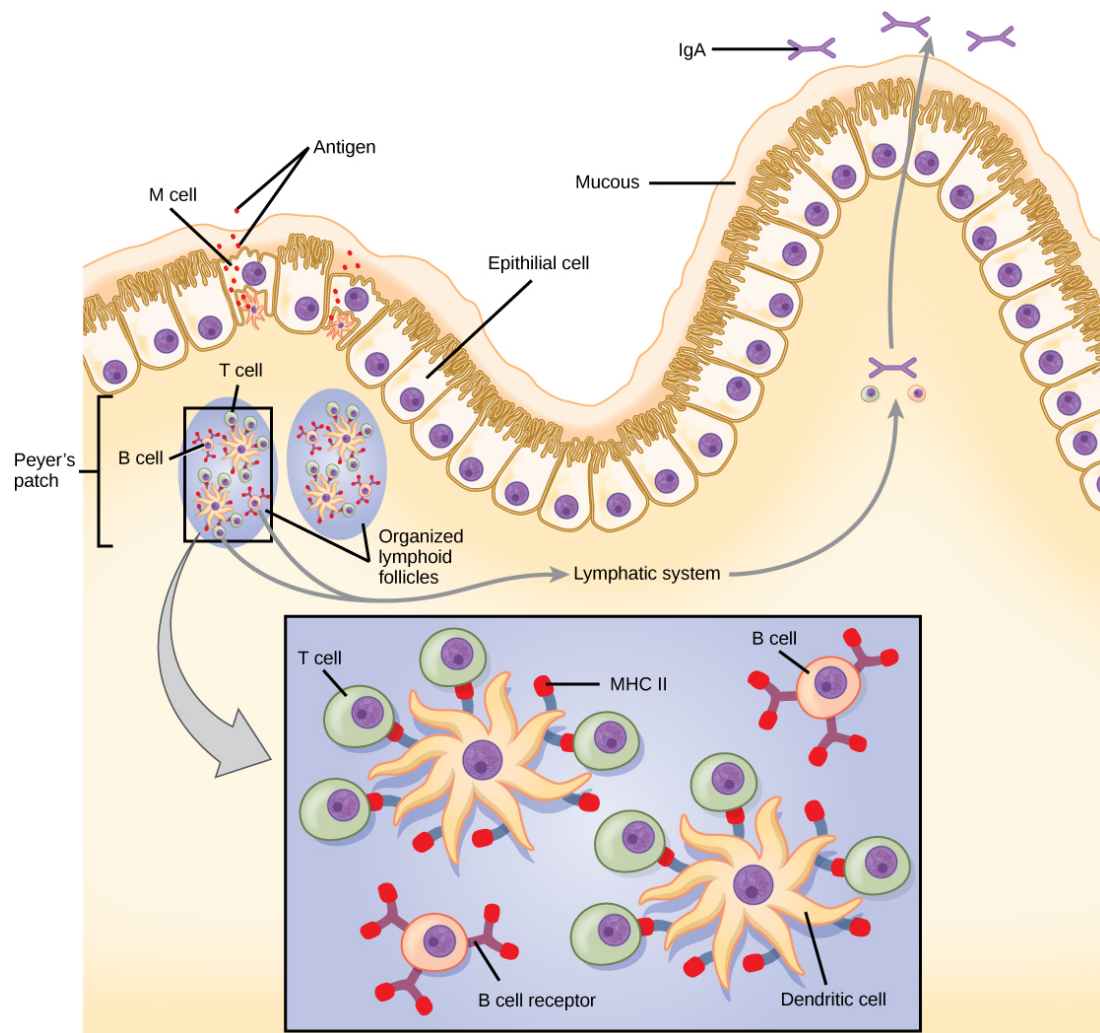
Natural killer (NK) cells recognize the MHC I receptor on healthy cells. If MHC I is absent, the cell is lysed.

Based on what you know about MHC receptors, why do you think an organ transplanted from an incompatible donor to a recipient will be rejected?

Plasma cells and CTLs are collectively called **effector cells**: they represent differentiated versions of their naïve counterparts, and they are involved in bringing about the immune defense of killing pathogens and infected host cells.

## Mucosal Surfaces and Immune Tolerance

The innate and adaptive immune responses discussed thus far comprise the systemic immune system (affecting the whole body), which is distinct from the mucosal immune system. Mucosal immunity is formed by mucosa-associated lymphoid tissue, which functions independently of the systemic immune system, and which has its own innate and adaptive components. **Mucosa-associated lymphoid tissue (MALT)**, illustrated in [\[link\]](#), is a collection of lymphatic tissue that combines with epithelial tissue lining the mucosa throughout the body. This tissue functions as the immune barrier and response in areas of the body with direct contact to the external environment. The systemic and mucosal immune systems use many of the same cell types. Foreign particles that make their way to MALT are taken up by absorptive epithelial cells called M cells and delivered to APCs located directly below the mucosal tissue. M cells function in the transport described, and are located in the Peyer's patch, a lymphoid nodule. APCs of the mucosal immune system are primarily dendritic cells, with B cells and macrophages having minor roles. Processed antigens displayed on APCs are detected by T cells in the MALT and at various mucosal induction sites, such as the tonsils, adenoids, appendix, or the mesenteric lymph nodes of the intestine. Activated T cells then migrate through the lymphatic system and into the circulatory system to mucosal sites of infection.



The topology and function of intestinal MALT is shown. Pathogens are taken up by M cells in the intestinal epithelium and excreted into a pocket formed by the inner surface of the cell. The pocket contains antigen-presenting cells such as dendritic cells, which engulf the antigens, then present them with MHC II molecules on the cell surface. The dendritic cells migrate to an underlying tissue called a Peyer's patch. Antigen-presenting cells, T cells, and B cells aggregate within the Peyer's patch, forming organized lymphoid follicles. There, some T cells and B cells are activated. Other antigen-loaded dendritic cells migrate through the lymphatic system where they activate B cells, T cells, and plasma cells in the lymph nodes. The activated cells then return to MALT tissue effector

sites. IgA and other antibodies are secreted into the intestinal lumen.

MALT is a crucial component of a functional immune system because mucosal surfaces, such as the nasal passages, are the first tissues onto which inhaled or ingested pathogens are deposited. The mucosal tissue includes the mouth, pharynx, and esophagus, and the gastrointestinal, respiratory, and urogenital tracts.

The immune system has to be regulated to prevent wasteful, unnecessary responses to harmless substances, and more importantly so that it does not attack “self.” The acquired ability to prevent an unnecessary or harmful immune response to a detected foreign substance known not to cause disease is described as **immune tolerance**. Immune tolerance is crucial for maintaining mucosal homeostasis given the tremendous number of foreign substances (such as food proteins) that APCs of the oral cavity, pharynx, and gastrointestinal mucosa encounter. Immune tolerance is brought about by specialized APCs in the liver, lymph nodes, small intestine, and lung that present harmless antigens to an exceptionally diverse population of **regulatory T (T<sub>reg</sub>) cells**, specialized lymphocytes that suppress local inflammation and inhibit the secretion of stimulatory immune factors. The combined result of T<sub>reg</sub> cells is to prevent immunologic activation and inflammation in undesired tissue compartments and to allow the immune system to focus on pathogens instead. In addition to promoting immune tolerance of harmless antigens, other subsets of T<sub>reg</sub> cells are involved in the prevention of the **autoimmune response**, which is an inappropriate immune response to host cells or self-antigens. Another T<sub>reg</sub> class suppresses immune responses to harmful pathogens after the infection has cleared to minimize host cell damage induced by inflammation and cell lysis.

## Immunological Memory

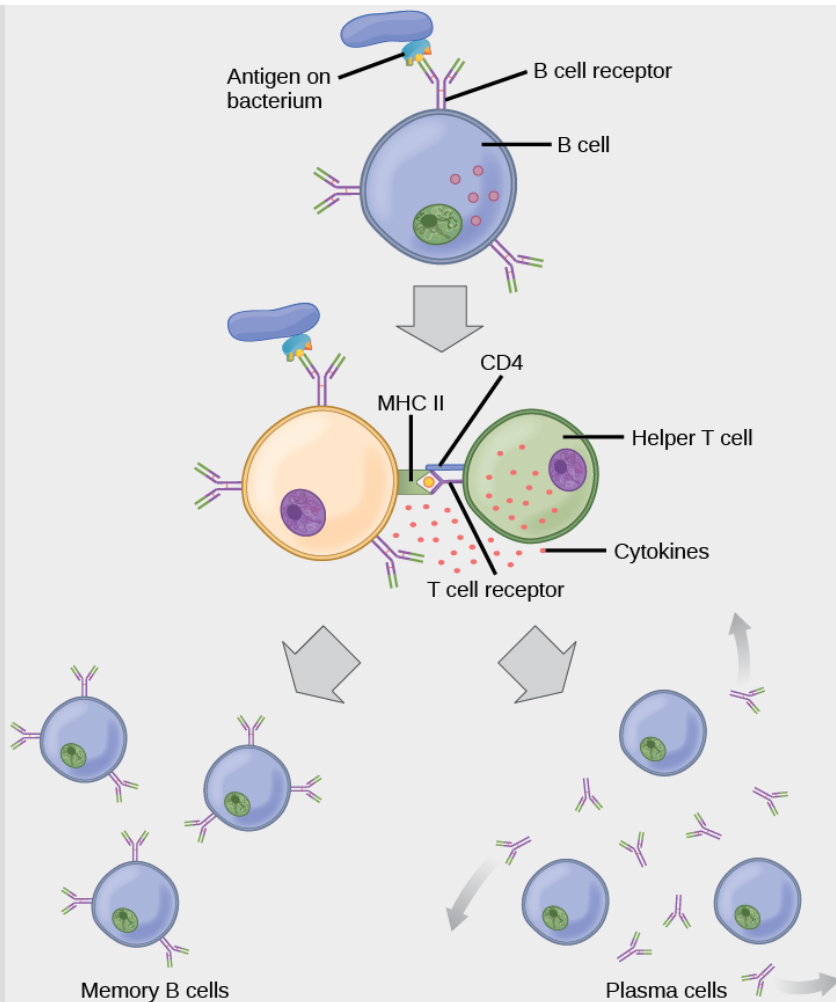
The adaptive immune system possesses a memory component that allows for an efficient and dramatic response upon reinvasion of the same pathogen. Memory is handled by the adaptive immune system with little

reliance on cues from the innate response. During the adaptive immune response to a pathogen that has not been encountered before, called a primary response, plasma cells secreting antibodies and differentiated T cells increase, then plateau over time. As B and T cells mature into effector cells, a subset of the naïve populations differentiates into B and T memory cells with the same antigen specificities, as illustrated in [\[link\]](#).

A **memory cell** is an antigen-specific B or T lymphocyte that does not differentiate into effector cells during the primary immune response, but that can immediately become effector cells upon re-exposure to the same pathogen. During the primary immune response, memory cells do not respond to antigens and do not contribute to host defenses. As the infection is cleared and pathogenic stimuli subside, the effectors are no longer needed, and they undergo apoptosis. In contrast, the memory cells persist in the circulation.

**Note:**

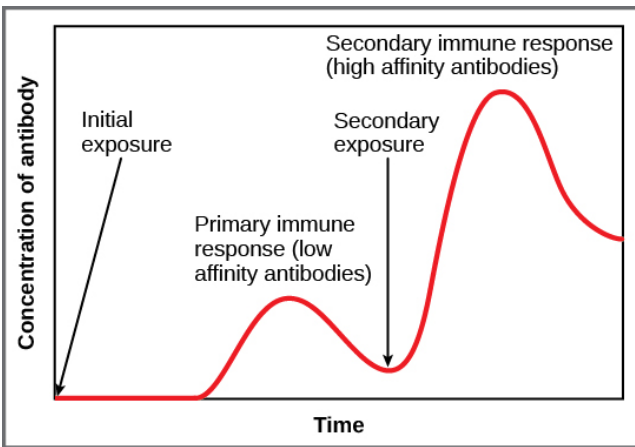
Art Connection



After initially binding an antigen to the B cell receptor (BCR), a B cell internalizes the antigen and presents it on MHC II. A helper T cell recognizes the MHC II–antigen complex and activates the B cell. As a result, memory B cells and plasma cells are made.

The Rh antigen is found on Rh-positive red blood cells. An Rh-negative female can usually carry an Rh-positive fetus to term without difficulty. However, if she has a second Rh-positive fetus, her body may launch an immune attack that causes hemolytic disease of the newborn. Why do you think hemolytic disease is only a problem during the second or subsequent pregnancies?

If the pathogen is never encountered again during the individual's lifetime, B and T memory cells will circulate for a few years or even several decades and will gradually die off, having never functioned as effector cells. However, if the host is re-exposed to the same pathogen type, circulating memory cells will immediately differentiate into plasma cells and CTLs without input from APCs or  $T_H$  cells. One reason the adaptive immune response is delayed is because it takes time for naïve B and T cells with the appropriate antigen specificities to be identified and activated. Upon reinfection, this step is skipped, and the result is a more rapid production of immune defenses. Memory B cells that differentiate into plasma cells output tens to hundreds-fold greater antibody amounts than were secreted during the primary response, as the graph in [\[link\]](#) illustrates. This rapid and dramatic antibody response may stop the infection before it can even become established, and the individual may not realize they had been exposed.



In the primary response to infection, antibodies are secreted first from plasma cells. Upon re-exposure to the same pathogen, memory cells differentiate into antibody-secreting plasma cells that output a greater amount of antibody for a longer period of time.



Vaccination is based on the knowledge that exposure to noninfectious antigens, derived from known pathogens, generates a mild primary immune response. The immune response to vaccination may not be perceived by the host as illness but still confers immune memory. When exposed to the corresponding pathogen to which an individual was vaccinated, the reaction is similar to a secondary exposure. Because each reinfection generates more memory cells and increased resistance to the pathogen, and because some memory cells die, certain vaccine courses involve one or more booster vaccinations to mimic repeat exposures: for instance, tetanus boosters are necessary every ten years because the memory cells only live that long.

### **Mucosal Immune Memory**

A subset of T and B cells of the mucosal immune system differentiates into memory cells just as in the systemic immune system. Upon reinvasion of the same pathogen type, a pronounced immune response occurs at the mucosal site where the original pathogen deposited, but a collective defense is also organized within interconnected or adjacent mucosal tissue. For instance, the immune memory of an infection in the oral cavity would also elicit a response in the pharynx if the oral cavity was exposed to the same pathogen.

#### **Note:**

##### **Career Connection**

##### **Vaccinologist**

Vaccination (or immunization) involves the delivery, usually by injection as shown in [\[link\]](#), of noninfectious antigen(s) derived from known pathogens. Other components, called adjuvants, are delivered in parallel to help stimulate the immune response. Immunological memory is the reason vaccines work. Ideally, the effect of vaccination is to elicit immunological

memory, and thus resistance to specific pathogens without the individual having to experience an infection.



Vaccines are often delivered by injection into the arm. (credit: U.S. Navy Photographer's Mate Airman Apprentice Christopher D. Blachly)

Vaccinologists are involved in the process of vaccine development from the initial idea to the availability of the completed vaccine. This process can take decades, can cost millions of dollars, and can involve many obstacles along the way. For instance, injected vaccines stimulate the systemic immune system, eliciting humoral and cell-mediated immunity, but have little effect on the mucosal response, which presents a challenge because many pathogens are deposited and replicate in mucosal compartments, and the injection does not provide the most efficient immune memory for these disease agents. For this reason, vaccinologists are actively involved in developing new vaccines that are applied via intranasal, aerosol, oral, or transcutaneous (absorbed through the skin) delivery methods. Importantly, mucosal-administered vaccines elicit both mucosal and systemic immunity and produce the same level of disease resistance as injected vaccines.



The polio vaccine  
can be administered  
orally. (credit:  
modification of  
work by UNICEF  
Sverige)

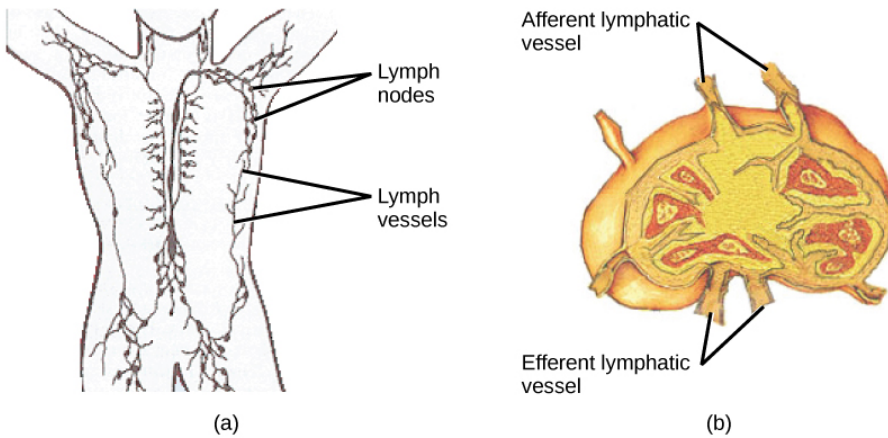
Currently, a version of intranasal influenza vaccine is available, and the polio and typhoid vaccines can be administered orally, as shown in [\[link\]](#). Similarly, the measles and rubella vaccines are being adapted to aerosol delivery using inhalation devices. Eventually, transgenic plants may be engineered to produce vaccine antigens that can be eaten to confer disease resistance. Other vaccines may be adapted to rectal or vaginal application to elicit immune responses in rectal, genitourinary, or reproductive mucosa. Finally, vaccine antigens may be adapted to transdermal application in which the skin is lightly scraped and microneedles are used to pierce the outermost layer. In addition to mobilizing the mucosal immune response, this new generation of vaccines may end the anxiety associated with injections and, in turn, improve patient participation.

## Primary Centers of the Immune System

Although the immune system is characterized by circulating cells throughout the body, the regulation, maturation, and intercommunication of immune factors occur at specific sites. The blood circulates immune cells, proteins, and other factors through the body. Approximately 0.1 percent of all cells in the blood are leukocytes, which encompass monocytes (the precursor of macrophages) and lymphocytes. The majority of cells in the blood are erythrocytes (red blood cells). **Lymph** is a watery fluid that bathes tissues and organs with protective white blood cells and does not contain erythrocytes. Cells of the immune system can travel between the distinct lymphatic and blood circulatory systems, which are separated by interstitial space, by a process called extravasation (passing through to surrounding tissue).

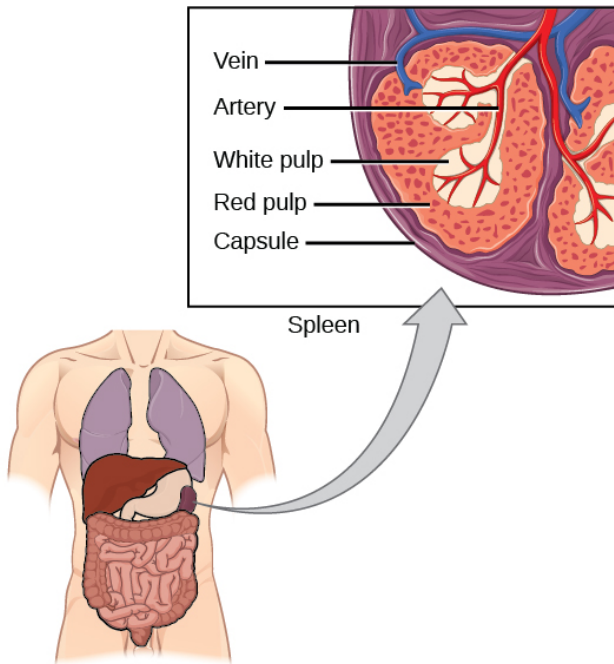
The cells of the immune system originate from hematopoietic stem cells in the bone marrow. Cytokines stimulate these stem cells to differentiate into immune cells. B cell maturation occurs in the bone marrow, whereas naïve T cells transit from the bone marrow to the thymus for maturation. In the thymus, immature T cells that express TCRs complementary to self-antigens are destroyed. This process helps prevent autoimmune responses.

On maturation, T and B lymphocytes circulate to various destinations. Lymph nodes scattered throughout the body, as illustrated in [\[link\]](#), house large populations of T and B cells, dendritic cells, and macrophages. Lymph gathers antigens as it drains from tissues. These antigens then are filtered through lymph nodes before the lymph is returned to circulation. APCs in the lymph nodes capture and process antigens and inform nearby lymphocytes about potential pathogens.



(a) Lymphatic vessels carry a clear fluid called lymph throughout the body. The liquid enters (b) lymph nodes through afferent vessels. Lymph nodes are filled with lymphocytes that purge infecting cells. The lymph then exits through efferent vessels. (credit: modification of work by NIH, NCI)

The spleen houses B and T cells, macrophages, dendritic cells, and NK cells. The spleen, shown in [\[link\]](#), is the site where APCs that have trapped foreign particles in the blood can communicate with lymphocytes. Antibodies are synthesized and secreted by activated plasma cells in the spleen, and the spleen filters foreign substances and antibody-complexed pathogens from the blood. Functionally, the spleen is to the blood as lymph nodes are to the lymph.



The spleen is similar to a lymph node but is much larger and filters blood instead of lymph. Blood enters the spleen through arteries and exits through veins. The spleen contains two types of tissue: red pulp and white pulp. Red pulp consists of cavities that store blood. Within the red pulp, damaged red blood cells are removed and replaced by new ones. White pulp is rich in lymphocytes that remove antigen-coated bacteria from the blood.  
(credit: modification of work by NCI)

## Section Summary

The adaptive immune response is a slower-acting, longer-lasting, and more specific response than the innate response. However, the adaptive response requires information from the innate immune system to function. APCs display antigens via MHC molecules to complementary naïve T cells. In response, the T cells differentiate and proliferate, becoming T<sub>H</sub> cells or CTLs. T<sub>H</sub> cells stimulate B cells that have engulfed and presented pathogen-derived antigens. B cells differentiate into plasma cells that secrete antibodies, whereas CTLs induce apoptosis in intracellularly infected or cancerous cells. Memory cells persist after a primary exposure to a pathogen. If re-exposure occurs, memory cells differentiate into effector cells without input from the innate immune system. The mucosal immune system is largely independent from the systemic immune system but functions in a parallel fashion to protect the extensive mucosal surfaces of the body.

## Art Connections

### Exercise:

#### Problem:

[\[link\]](#) Which of the following statements about T cells is false?

- a. Helper T cells release cytokines while cytotoxic T cells kill the infected cell.
- b. Helper T cells are CD4<sup>+</sup>, while cytotoxic T cells are CD8<sup>+</sup>.
- c. MHC II is a receptor found on most body cells, while MHC I is a receptor found on immune cells only.
- d. The T cell receptor is found on both CD4<sup>+</sup> and CD8<sup>+</sup> T cells.

---

#### Solution:

[\[link\]](#) C

### Exercise:

**Problem:**

[\[link\]](#) Based on what you know about MHC receptors, why do you think an organ transplanted from an incompatible donor to a recipient will be rejected?

---

**Solution:**

[\[link\]](#) MHC receptors differ from person to person. Thus, MHC receptors on an incompatible donor are considered “non-self” and are rejected by the immune system.

**Exercise:****Problem:**

[\[link\]](#) The Rh antigen is found on Rh-positive red blood cells. An Rh-negative female can usually carry an Rh-positive fetus to term without difficulty. However, if she has a second Rh-positive fetus, her body may launch an immune attack that causes hemolytic disease of the newborn. Why do you think hemolytic disease is only a problem during the second or subsequent pregnancies?

---

**Solution:**

[\[link\]](#) If the blood of the mother and fetus mixes, memory cells that recognize the Rh antigen can form late in the first pregnancy. During subsequent pregnancies, these memory cells launch an immune attack on the fetal blood cells. Injection of anti-Rh antibody during the first pregnancy prevents the immune response from occurring.

**Review Questions****Exercise:**



**Problem:**

Which of the following is both a phagocyte and an antigen-presenting cell?

- a. NK cell
- b. eosinophil
- c. neutrophil
- d. macrophage

---

**Solution:**

D

**Exercise:****Problem:**

Which immune cells bind MHC molecules on APCs via CD8 coreceptors on their cell surfaces?

- a. T<sub>H</sub> cells
- b. CTLs
- c. mast cells
- d. basophils

---

**Solution:**

B

**Exercise:**

**Problem:** What “self” pattern is identified by NK cells?

- a. altered self
- b. missing self
- c. normal self

d. non-self

---

**Solution:**

B

**Exercise:**

**Problem:**

The acquired ability to prevent an unnecessary or destructive immune reaction to a harmless foreign particle, such as a food protein, is called \_\_\_\_\_.

- a. the  $T_H2$  response
- b. allergy
- c. immune tolerance
- d. autoimmunity

---

**Solution:**

C

**Exercise:**

**Problem:**

A memory B cell can differentiate upon re-exposure to a pathogen of which cell type?

- a. CTL
- b. naïve B cell
- c. memory T cell
- d. plasma cell

---

**Solution:**

D

**Exercise:****Problem:**

Foreign particles circulating in the blood are filtered by the \_\_\_\_\_.

- a. spleen
- b. lymph nodes
- c. MALT
- d. lymph

---

**Solution:**

A

**Free Response****Exercise:**

**Problem:** Explain the difference between an epitope and an antigen.

---

**Solution:**

An antigen is a molecule that reacts with some component of the immune response (antibody, B cell receptor, T cell receptor). An epitope is the region on the antigen through which binding with the immune component actually occurs.

**Exercise:**

**Problem:** What is a naïve B or T cell?

---

**Solution:**

A naïve T or B cell is one that has not been activated by binding to the appropriate epitope. Naïve T and B cells cannot produce responses.

**Exercise:**

**Problem:** How does the  $T_H1$  response differ from the  $T_H2$  response?

---

**Solution:**

The  $T_H1$  response involves the secretion of cytokines to stimulate macrophages and CTLs and improve their destruction of intracellular pathogens and tumor cells. It is associated with inflammation. The  $T_H2$  response is involved in the stimulation of B cells into plasma cells that synthesize and secrete antibodies.

**Exercise:****Problem:**

In mammalian adaptive immune systems, T cell receptors are extraordinarily diverse. What function of the immune system results from this diversity, and how is this diversity achieved?

---

**Solution:**

The diversity of TCRs allows the immune system to have millions of different T cells, and thereby to be specific in distinguishing antigens. This diversity arises from mutation and recombination in the genes that encode the variable regions of TCRs.

**Exercise:****Problem:**

How do B and T cells differ with respect to antigens that they bind?

---

**Solution:**

T cells bind antigens that have been digested and embedded in MHC molecules by APCs. In contrast, B cells function themselves as APCs to bind intact, unprocessed antigens.

**Exercise:**

**Problem:**

Why is the immune response after reinfection much faster than the adaptive immune response after the initial infection?

---

**Solution:**

Upon reinfection, the memory cells will immediately differentiate into plasma cells and CTLs without input from APCs or  $T_H$  cells. In contrast, the adaptive immune response to the initial infection requires time for naïve B and T cells with the appropriate antigen specificities to be identified and activated.

**Glossary**

adaptive immunity

immunity that has memory and occurs after exposure to an antigen either from a pathogen or a vaccination

antigen

foreign or “non-self” protein that triggers the immune response

antigen-presenting cell (APC)

immune cell that detects, engulfs, and informs the adaptive immune response about an infection by presenting the processed antigen on the cell surface

autoimmune response

inappropriate immune response to host cells or self-antigens

cell-mediated immune response

adaptive immune response that is carried out by T cells

clonal selection

activation of B cells corresponding to one specific BCR variant and the dramatic proliferation of that variant

cytotoxic T lymphocyte (CTL)

adaptive immune cell that directly kills infected cells via perforin and granzymes, and releases cytokines to enhance the immune response

dendritic cell

immune cell that processes antigen material and presents it on the surface of other cells to induce an immune response

effector cell

lymphocyte that has differentiated, such as a B cell, plasma cell, or cytotoxic T lymphocyte

epitope

small component of an antigen that is specifically recognized by antibodies, B cells, and T cells; the antigenic determinant

helper T lymphocyte ( $T_H$ )

cell of the adaptive immune system that binds APCs via MHC II molecules and stimulates B cells or secretes cytokines to initiate the immune response

humoral immune response

adaptive immune response that is controlled by activated B cells and antibodies

immune tolerance

acquired ability to prevent an unnecessary or harmful immune response to a detected foreign body known not to cause disease or to self-antigens

lymph

watery fluid that bathes tissues and organs with protective white blood cells and does not contain erythrocytes

mucosa-associated lymphoid tissue (MALT)

collection of lymphatic tissue that combines with epithelial tissue lining the mucosa throughout the body

memory cell

antigen-specific B or T lymphocyte that does not differentiate into effector cells during the primary immune response but that can immediately become an effector cell upon re-exposure to the same pathogen

plasma cell

immune cell that secretes antibodies; these cells arise from B cells that were stimulated by antigens

regulatory T (T<sub>reg</sub>) cell

specialized lymphocyte that suppresses local inflammation and inhibits the secretion of cytokines, antibodies, and other stimulatory immune factors; involved in immune tolerance

## Antibodies

By the end of this section, you will be able to:

- Explain cross-reactivity
- Describe the structure and function of antibodies
- Discuss antibody production

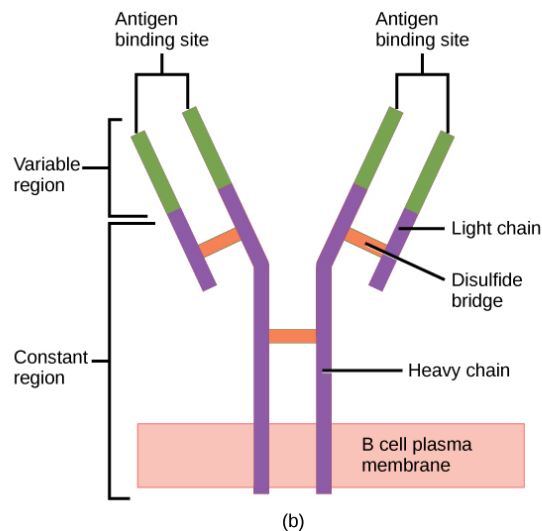
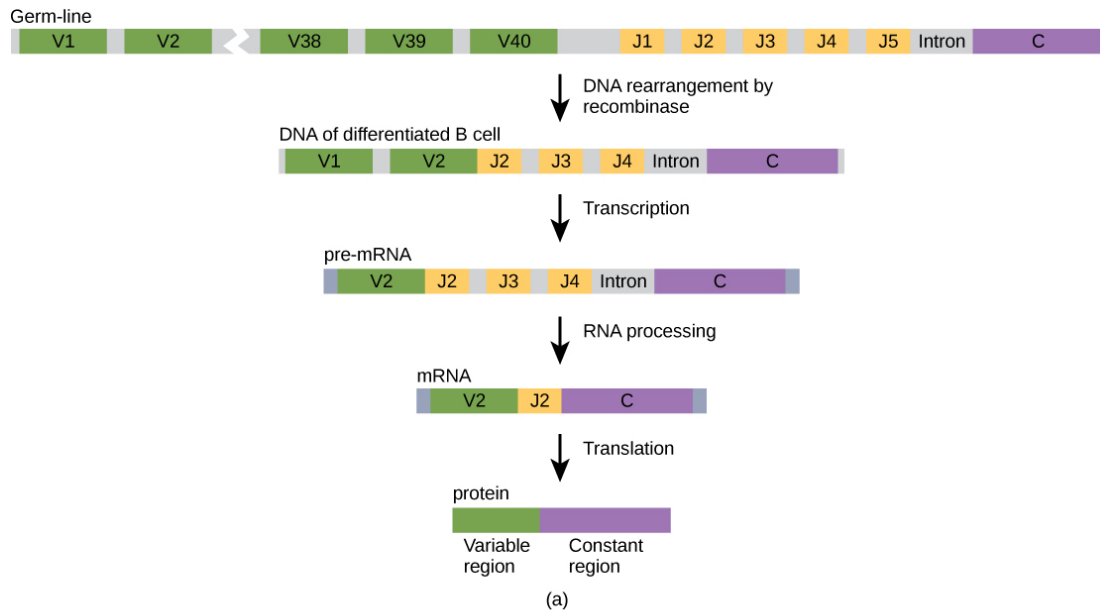
An **antibody**, also known as an immunoglobulin (Ig), is a protein that is produced by plasma cells after stimulation by an antigen. Antibodies are the functional basis of humoral immunity. Antibodies occur in the blood, in gastric and mucus secretions, and in breast milk. Antibodies in these bodily fluids can bind pathogens and mark them for destruction by phagocytes before they can infect cells.

## Antibody Structure

An antibody molecule is comprised of four polypeptides: two identical heavy chains (large peptide units) that are partially bound to each other in a “Y” formation, which are flanked by two identical light chains (small peptide units), as illustrated in [\[link\]](#). Bonds between the cysteine amino acids in the antibody molecule attach the polypeptides to each other. The areas where the antigen is recognized on the antibody are variable domains and the antibody base is composed of constant domains.

In germ-line B cells, the variable region of the light chain gene has 40 variable (V) and five joining (J) segments. An enzyme called DNA recombinase randomly excises most of these segments out of the gene, and splices one V segment to one J segment. During RNA processing, all but one V and J segment are spliced out. Recombination and splicing may result in over  $10^6$  possible VJ combinations. As a result, each differentiated B cell in the human body typically has a unique variable chain. The constant domain, which does not bind antibody, is the same for all antibodies.





(a) As a germ-line B cell matures, an enzyme called DNA recombinase randomly excises V and J segments from the light chain gene. Splicing at the mRNA level results in further gene rearrangement. As a result, (b) each antibody has a unique variable region capable of binding a different antigen.






Similar to TCRs and BCRs, antibody diversity is produced by the mutation and recombination of approximately 300 different gene segments encoding

the light and heavy chain variable domains in precursor cells that are destined to become B cells. The variable domains from the heavy and light chains interact to form the binding site through which an antibody can bind a specific epitope on an antigen. The numbers of repeated constant domains in Ig classes are the same for all antibodies corresponding to a specific class. Antibodies are structurally similar to the extracellular component of the BCRs, and B cell maturation to plasma cells can be visualized in simple terms as the cell acquires the ability to secrete the extracellular portion of its BCR in large quantities.

## **Antibody Classes**

Antibodies can be divided into five classes—IgM, IgG, IgA, IgD, IgE—based on their physiochemical, structural, and immunological properties. IgGs, which make up about 80 percent of all antibodies, have heavy chains that consist of one variable domain and three identical constant domains. IgA and IgD also have three constant domains per heavy chain, whereas IgM and IgE each have four constant domains per heavy chain. The variable domain determines binding specificity and the constant domain of the heavy chain determines the immunological mechanism of action of the corresponding antibody class. It is possible for two antibodies to have the same binding specificities but be in different classes and, therefore, to be involved in different functions.

After an adaptive defense is produced against a pathogen, typically plasma cells first secrete IgM into the blood. BCRs on naïve B cells are of the IgM class and occasionally IgD class. IgM molecules make up approximately ten percent of all antibodies. Prior to antibody secretion, plasma cells assemble IgM molecules into pentamers (five individual antibodies) linked by a joining (J) chain, as shown in [\[link\]](#). The pentamer arrangement means that these macromolecules can bind ten identical antigens. However, IgM molecules released early in the adaptive immune response do not bind to antigens as stably as IgGs, which are one of the possible types of antibodies secreted in large quantities upon re-exposure to the same pathogen. [\[link\]](#) summarizes the properties of immunoglobulins and illustrates their basic structures.

Name	Properties	Structure
IgA	Found in mucous, saliva, tears, and breast milk. Protects against pathogens.	
IgD	Part of the B cell receptor. Activates basophils and mast cells.	
IgE	Protects against parasitic worms. Responsible for allergic reactions.	
IgG	Secreted by plasma cells in the blood. Able to cross the placenta into the fetus.	
IgM	May be attached to the surface of a B cell or secreted into the blood. Responsible for early stages of immunity.	

Immunoglobulins have different functions, but all are composed of light and heavy chains that form a Y-shaped structure.

IgAs populate the saliva, tears, breast milk, and mucus secretions of the gastrointestinal, respiratory, and genitourinary tracts. Collectively, these bodily fluids coat and protect the extensive mucosa (4000 square feet in humans). The total number of IgA molecules in these bodily secretions is greater than the number of IgG molecules in the blood serum. A small amount of IgA is also secreted into the serum in monomeric form. Conversely, some IgM is secreted into bodily fluids of the mucosa. Similar to IgM, IgA molecules are secreted as polymeric structures linked with a J chain. However, IgAs are secreted mostly as dimeric molecules, not pentamers.

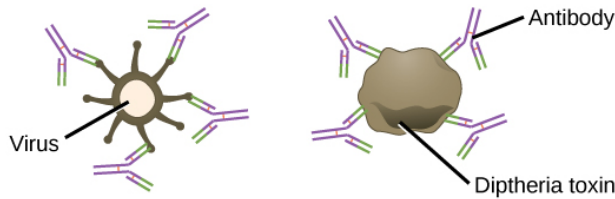
IgE is present in the serum in small quantities and is best characterized in its role as an allergy mediator. IgD is also present in small quantities. Similar to IgM, BCRs of the IgD class are found on the surface of naïve B cells. This class supports antigen recognition and maturation of B cells to plasma cells.

## Antibody Functions

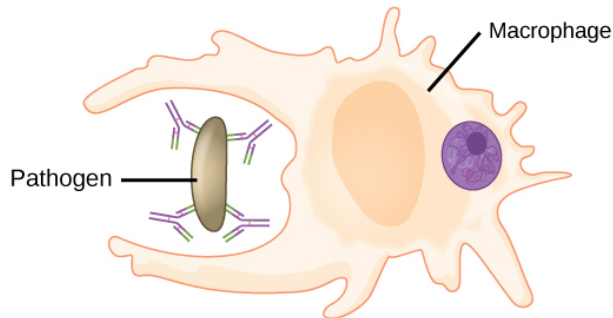
Differentiated plasma cells are crucial players in the humoral response, and the antibodies they secrete are particularly significant against extracellular pathogens and toxins. Antibodies circulate freely and act independently of plasma cells. Antibodies can be transferred from one individual to another to temporarily protect against infectious disease. For instance, a person who has recently produced a successful immune response against a particular disease agent can donate blood to a nonimmune recipient and confer temporary immunity through antibodies in the donor's blood serum. This phenomenon is called **passive immunity**; it also occurs naturally during breastfeeding, which makes breastfed infants highly resistant to infections during the first few months of life.

Antibodies coat extracellular pathogens and neutralize them, as illustrated in [\[link\]](#), by blocking key sites on the pathogen that enhance their infectivity (such as receptors that “dock” pathogens on host cells). Antibody neutralization can prevent pathogens from entering and infecting host cells, as opposed to the CTL-mediated approach of killing cells that are already infected to prevent progression of an established infection. The neutralized antibody-coated pathogens can then be filtered by the spleen and eliminated in urine or feces.

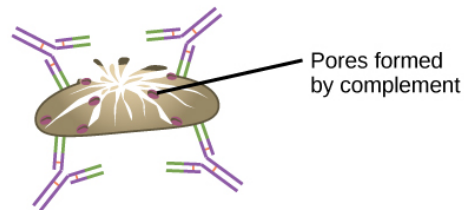
**(a) Neutralization** Antibodies prevent a virus or toxic protein from binding their target.



**(b) Opsonization** A pathogen tagged by antibodies is consumed by a macrophage or neutrophil.



**(c) Complement activation** Antibodies attached to the surface of a pathogen cell activate the complement system.



Antibodies may inhibit infection by (a) preventing the antigen from binding its target, (b) tagging a pathogen for destruction by macrophages or neutrophils, or (c) activating the complement cascade.

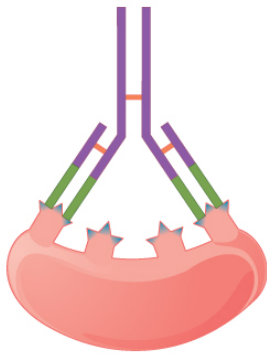
Antibodies also mark pathogens for destruction by phagocytic cells, such as macrophages or neutrophils, because phagocytic cells are highly attracted to macromolecules complexed with antibodies. Phagocytic enhancement by antibodies is called opsonization. In a process called complement fixation,

IgM and IgG in serum bind to antigens and provide docking sites onto which sequential complement proteins can bind. The combination of antibodies and complement enhances opsonization even further and promotes rapid clearing of pathogens.

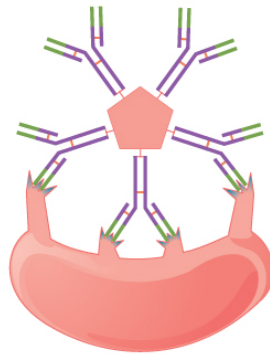
### **Affinity, Avidity, and Cross Reactivity**

Not all antibodies bind with the same strength, specificity, and stability. In fact, antibodies exhibit different **affinities** (attraction) depending on the molecular complementarity between antigen and antibody molecules, as illustrated in [\[link\]](#). An antibody with a higher affinity for a particular antigen would bind more strongly and stably, and thus would be expected to present a more challenging defense against the pathogen corresponding to the specific antigen.

(a) Affinity versus avidity

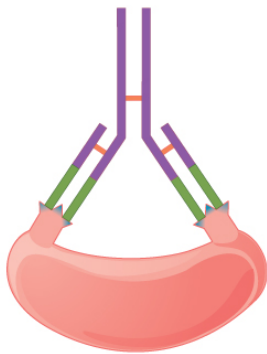


Affinity refers to the strength of a single antibody–antigen interaction. Each IgG antigen binding site typically has high affinity for its target.

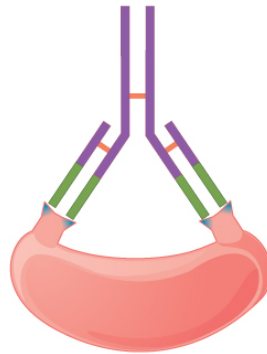


Avidity refers to the strength of all interactions combined. IgM typically has low affinity antigen binding sites, but there are ten of them, so avidity is high.

(b) Cross reactivity



An antibody may react with two different epitopes.



(a) Affinity refers to the strength of single interaction between antigen and antibody, while avidity refers to the strength of all interactions combined. (b) An antibody may cross react with different epitopes.

The term **avidity** describes binding by antibody classes that are secreted as joined, multivalent structures (such as IgM and IgA). Although avidity measures the strength of binding, just as affinity does, the avidity is not simply the sum of the affinities of the antibodies in a multimeric structure.

The avidity depends on the number of identical binding sites on the antigen being detected, as well as other physical and chemical factors. Typically, multimeric antibodies, such as pentameric IgM, are classified as having lower affinity than monomeric antibodies, but high avidity. Essentially, the fact that multimeric antibodies can bind many antigens simultaneously balances their slightly lower binding strength for each antibody/antigen interaction.

Antibodies secreted after binding to one epitope on an antigen may exhibit cross reactivity for the same or similar epitopes on different antigens. Because an epitope corresponds to such a small region (the surface area of about four to six amino acids), it is possible for different macromolecules to exhibit the same molecular identities and orientations over short regions. **Cross reactivity** describes when an antibody binds not to the antigen that elicited its synthesis and secretion, but to a different antigen.

Cross reactivity can be beneficial if an individual develops immunity to several related pathogens despite having only been exposed to or vaccinated against one of them. For instance, antibody cross reactivity may occur against the similar surface structures of various Gram-negative bacteria. Conversely, antibodies raised against pathogenic molecular components that resemble self molecules may incorrectly mark host cells for destruction and cause autoimmune damage. Patients who develop systemic lupus erythematosus (SLE) commonly exhibit antibodies that react with their own DNA. These antibodies may have been initially raised against the nucleic acid of microorganisms but later cross-reacted with self-antigens. This phenomenon is also called molecular mimicry.

## **Antibodies of the Mucosal Immune System**

Antibodies synthesized by the mucosal immune system include IgA and IgM. Activated B cells differentiate into mucosal plasma cells that synthesize and secrete dimeric IgA, and to a lesser extent, pentameric IgM. Secreted IgA is abundant in tears, saliva, breast milk, and in secretions of the gastrointestinal and respiratory tracts. Antibody secretion results in a local humoral response at epithelial surfaces and prevents infection of the mucosa by binding and neutralizing pathogens.



## Section Summary

Antibodies (immunoglobulins) are the molecules secreted from plasma cells that mediate the humoral immune response. There are five antibody classes; an antibody's class determines its mechanism of action and production site but does not control its binding specificity. Antibodies bind antigens via variable domains and can either neutralize pathogens or mark them for phagocytosis or activate the complement cascade.

## Review Questions

### Exercise:

#### Problem:

The structure of an antibody is similar to the extracellular component of which receptor?

- a. MHC I
- b. MHC II
- c. BCR
- d. none of the above

---

#### Solution:

C

### Exercise:

#### Problem:

The first antibody class to appear in the serum in response to a newly encountered pathogen is \_\_\_\_\_.

- a. IgM
- b. IgA
- c. IgG
- d. IgE

---

**Solution:**

A

**Exercise:**

**Problem:**

What is the most abundant antibody class detected in the serum upon reexposure to a pathogen or in reaction to a vaccine?

- a. IgM
- b. IgA
- c. IgG
- d. IgE

---

**Solution:**

C

**Exercise:**

**Problem:**

Breastfed infants typically are resistant to disease because of \_\_\_\_\_.

- a. active immunity
- b. passive immunity
- c. immune tolerance
- d. immune memory

---

**Solution:**

B

## Free Response

### Exercise:

**Problem:** What are the benefits and costs of antibody cross reactivity?

---

### Solution:

Cross reactivity of antibodies can be beneficial when it allows an individual's immune system to respond to an array of similar pathogens after being exposed to just one of them. A potential cost of cross reactivity is an antibody response to parts of the body (self) in addition to the appropriate antigen.

## Glossary

### affinity

attraction of molecular complementarity between antigen and antibody molecules

### antibody

protein that is produced by plasma cells after stimulation by an antigen; also known as an immunoglobulin

### avidity

total binding strength of a multivalent antibody with antigen

### cross reactivity

binding of an antibody to an epitope corresponding to an antigen that is different from the one the antibody was raised against

### passive immunity

transfer of antibodies from one individual to another to provide temporary protection against pathogens

## Disruptions in the Immune System

By the end of this section, you will be able to:

- Describe hypersensitivity
- Define autoimmunity

A functioning immune system is essential for survival, but even the sophisticated cellular and molecular defenses of the mammalian immune response can be defeated by pathogens at virtually every step. In the competition between immune protection and pathogen evasion, pathogens have the advantage of more rapid evolution because of their shorter generation time and other characteristics. For instance, *Streptococcus pneumoniae* (bacterium that cause pneumonia and meningitis) surrounds itself with a capsule that inhibits phagocytes from engulfing it and displaying antigens to the adaptive immune system. *Staphylococcus aureus* (bacterium that can cause skin infections, abscesses, and meningitis) synthesizes a toxin called leukocidin that kills phagocytes after they engulf the bacterium. Other pathogens can also hinder the adaptive immune system. HIV infects T<sub>H</sub> cells via their CD4 surface molecules, gradually depleting the number of T<sub>H</sub> cells in the body; this inhibits the adaptive immune system's capacity to generate sufficient responses to infection or tumors. As a result, HIV-infected individuals often suffer from infections that would not cause illness in people with healthy immune systems but which can cause devastating illness to immune-compromised individuals. Maladaptive responses of immune cells and molecules themselves can also disrupt the proper functioning of the entire system, leading to host cell damage that could become fatal.

## Immunodeficiency

Failures, insufficiencies, or delays at any level of the immune response can allow pathogens or tumor cells to gain a foothold and replicate or proliferate to high enough levels that the immune system becomes overwhelmed. **Immunodeficiency** is the failure, insufficiency, or delay in the response of the immune system, which may be acquired or inherited. Immunodeficiency can be acquired as a result of infection with certain pathogens (such as HIV), chemical exposure (including certain medical

treatments), malnutrition, or possibly by extreme stress. For instance, radiation exposure can destroy populations of lymphocytes and elevate an individual's susceptibility to infections and cancer. Dozens of genetic disorders result in immunodeficiencies, including Severe Combined Immunodeficiency (SCID), Bare lymphocyte syndrome, and MHC II deficiencies. Rarely, primary immunodeficiencies that are present from birth may occur. Neutropenia is one form in which the immune system produces a below-average number of neutrophils, the body's most abundant phagocytes. As a result, bacterial infections may go unrestricted in the blood, causing serious complications.

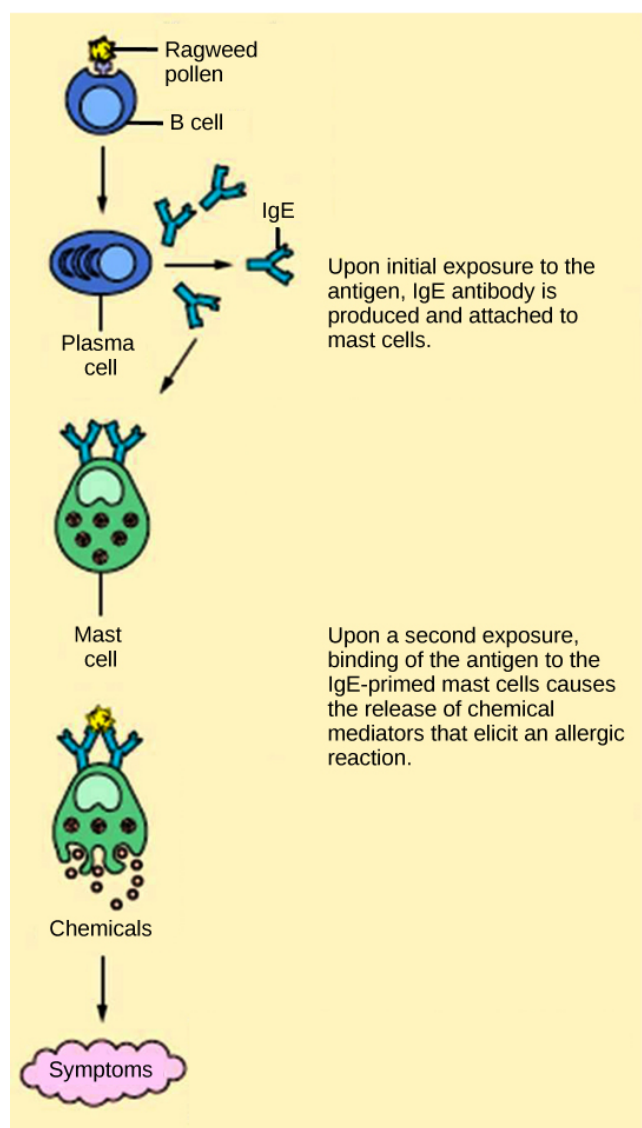
## **Hypersensitivities**

Maladaptive immune responses toward harmless foreign substances or self antigens that occur after tissue sensitization are termed **hypersensitivities**. The types of hypersensitivities include immediate, delayed, and autoimmunity. A large proportion of the population is affected by one or more types of hypersensitivity.

## **Allergies**

The immune reaction that results from immediate hypersensitivities in which an antibody-mediated immune response occurs within minutes of exposure to a harmless antigen is called an **allergy**. In the United States, 20 percent of the population exhibits symptoms of allergy or asthma, whereas 55 percent test positive against one or more allergens. Upon initial exposure to a potential allergen, an allergic individual synthesizes antibodies of the IgE class via the typical process of APCs presenting processed antigen to  $T_H$  cells that stimulate B cells to produce IgE. This class of antibodies also mediates the immune response to parasitic worms. The constant domain of the IgE molecules interact with mast cells embedded in connective tissues. This process primes, or sensitizes, the tissue. Upon subsequent exposure to the same allergen, IgE molecules on mast cells bind the antigen via their variable domains and stimulate the mast cell to release the modified amino acids histamine and serotonin; these chemical mediators then recruit

eosinophils which mediate allergic responses. [\[link\]](#) shows an example of an allergic response to ragweed pollen. The effects of an allergic reaction range from mild symptoms like sneezing and itchy, watery eyes to more severe or even life-threatening reactions involving intensely itchy welts or hives, airway contraction with severe respiratory distress, and plummeting blood pressure. This extreme reaction is known as anaphylactic shock. If not treated with epinephrine to counter the blood pressure and breathing effects, this condition can be fatal.



On first exposure to an allergen,

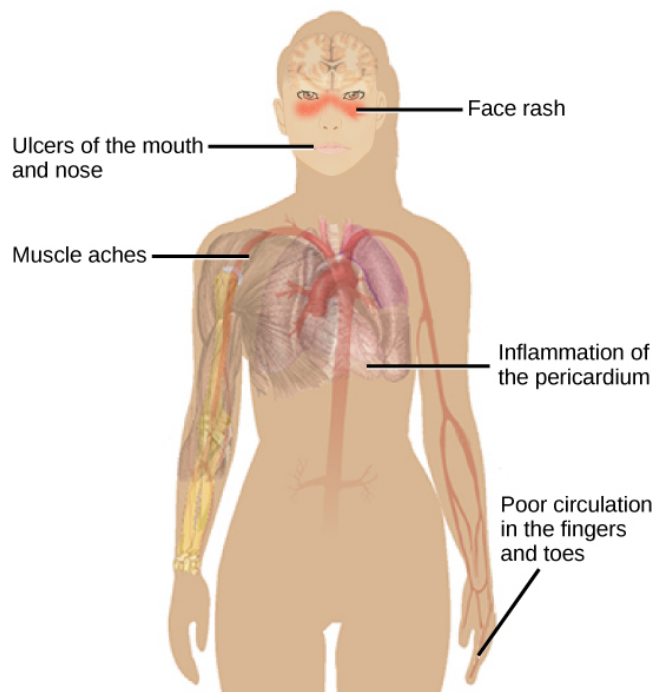
an IgE antibody is synthesized by plasma cells in response to a harmless antigen. The IgE molecules bind to mast cells, and on secondary exposure, the mast cells release histamines and other modulators that affect the symptoms of allergy. (credit: modification of work by NIH)

Delayed hypersensitivity is a cell-mediated immune response that takes approximately one to two days after secondary exposure for a maximal reaction to be observed. This type of hypersensitivity involves the  $T_H1$  cytokine-mediated inflammatory response and may manifest as local tissue lesions or contact dermatitis (rash or skin irritation). Delayed hypersensitivity occurs in some individuals in response to contact with certain types of jewelry or cosmetics. Delayed hypersensitivity facilitates the immune response to poison ivy and is also the reason why the skin test for tuberculosis results in a small region of inflammation on individuals who were previously exposed to *Mycobacterium tuberculosis*. That is also why cortisone is used to treat such responses: it will inhibit cytokine production.

## **Autoimmunity**

**Autoimmunity** is a type of hypersensitivity to self antigens that affects approximately five percent of the population. Most types of autoimmunity involve the humoral immune response. Antibodies that inappropriately mark self components as foreign are termed **autoantibodies**. In patients with the autoimmune disease myasthenia gravis, muscle cell receptors that induce contraction in response to acetylcholine are targeted by antibodies. The result is muscle weakness that may include marked difficulty with fine and/or gross motor functions. In systemic lupus erythematosus, a diffuse autoantibody response to the individual's own DNA and proteins results in

various systemic diseases. As illustrated in [\[link\]](#), systemic lupus erythematosus may affect the heart, joints, lungs, skin, kidneys, central nervous system, or other tissues, causing tissue damage via antibody binding, complement recruitment, lysis, and inflammation.



Systemic lupus erythematosus is characterized by autoimmunity to the individual's own DNA and/or proteins, which leads to varied dysfunction of the organs. (credit: modification of work by Mikael Häggström)

Autoimmunity can develop with time, and its causes may be rooted in molecular mimicry. Antibodies and TCRs may bind self antigens that are structurally similar to pathogen antigens, which the immune receptors first raised. As an example, infection with *Streptococcus pyogenes* (bacterium



that causes strep throat) may generate antibodies or T cells that react with heart muscle, which has a similar structure to the surface of *S. pyogenes*. These antibodies can damage heart muscle with autoimmune attacks, leading to rheumatic fever. Insulin-dependent (Type 1) diabetes mellitus arises from a destructive inflammatory T<sub>H</sub>1 response against insulin-producing cells of the pancreas. Patients with this autoimmunity must be injected with insulin that originates from other sources.

## Section Summary

Immune disruptions may involve insufficient immune responses or inappropriate immune targets. Immunodeficiency increases an individual's susceptibility to infections and cancers. Hypersensitivities are misdirected responses either to harmless foreign particles, as in the case of allergies, or to host factors, as in the case of autoimmunity. Reactions to self components may be the result of molecular mimicry.

## Review Questions

### Exercise:

**Problem:** Allergy to pollen is classified as:

- a. an autoimmune reaction
- b. immunodeficiency
- c. delayed hypersensitivity
- d. immediate hypersensitivity

---

### Solution:

D

### Exercise:

**Problem:** A potential cause of acquired autoimmunity is \_\_\_\_\_.

- a. tissue hypersensitivity
  - b. molecular mimicry
  - c. histamine release
  - d. radiation exposure
- 

**Solution:**

B

**Exercise:**

**Problem:** Autoantibodies are probably involved in:

- a. reactions to poison ivy
  - b. pollen allergies
  - c. systemic lupus erythematosus
  - d. HIV/AIDS
- 

**Solution:**

C

**Exercise:**

**Problem:** Which of the following diseases is not due to autoimmunity?

- a. rheumatic fever
  - b. systemic lupus erythematosus
  - c. diabetes mellitus
  - d. HIV/AIDS
- 

**Solution:**

D

## Glossary

### allergy

immune reaction that results from immediate hypersensitivities in which an antibody-mediated immune response occurs within minutes of exposure to a harmless antigen

### autoantibody

antibody that incorrectly marks “self” components as foreign and stimulates the immune response

### autoimmunity

type of hypersensitivity to self antigens

### hypersensitivities

spectrum of maladaptive immune responses toward harmless foreign particles or self antigens; occurs after tissue sensitization and includes immediate-type (allergy), delayed-type, and autoimmunity

### immunodeficiency

failure, insufficiency, or delay at any level of the immune system, which may be acquired or inherited

## Introduction

class="introduction"

Female seahorses  
produce eggs for  
reproduction that  
are then fertilized  
by the male.

Unlike almost all  
other animals, the  
male seahorse  
then gestates the  
young until birth.

(credit:  
modification of  
work by  
"cliff1066"/Flickr  
)



Animal reproduction is necessary for the survival of a species. In the animal kingdom, there are innumerable ways that species reproduce. Asexual reproduction produces genetically identical organisms (clones), whereas in sexual reproduction, the genetic material of two individuals combines to produce offspring that are genetically different from their parents. During sexual reproduction the male gamete (sperm) may be placed inside the female's body for internal fertilization, or the sperm and eggs may be released into the environment for external fertilization. Seahorses, like the one shown in [\[link\]](#), provide an example of the latter. Following a mating dance, the female lays eggs in the male seahorse's abdominal brood pouch where they are fertilized. The eggs hatch and the offspring develop in the pouch for several weeks.

## Reproduction Methods

By the end of this section, you will be able to:

- Describe advantages and disadvantages of asexual and sexual reproduction
- Discuss asexual reproduction methods
- Discuss sexual reproduction methods

Animals produce offspring through asexual and/or sexual reproduction. Both methods have advantages and disadvantages. **Asexual reproduction** produces offspring that are genetically identical to the parent because the offspring are all clones of the original parent. A single individual can produce offspring asexually and large numbers of offspring can be produced quickly. In a stable or predictable environment, asexual reproduction is an effective means of reproduction because all the offspring will be adapted to that environment. In an unstable or unpredictable environment asexually-reproducing species may be at a disadvantage because all the offspring are genetically identical and may not have the genetic variation to survive in new or different conditions. On the other hand, the rapid rates of asexual reproduction may allow for a speedy response to environmental changes if individuals have mutations. An additional advantage of asexual reproduction is that colonization of new habitats may be easier when an individual does not need to find a mate to reproduce.

During **sexual reproduction** the genetic material of two individuals is combined to produce genetically diverse offspring that differ from their parents. The genetic diversity of sexually produced offspring is thought to give species a better chance of surviving in an unpredictable or changing environment. Species that reproduce sexually must maintain two different types of individuals, males and females, which can limit the ability to colonize new habitats as both sexes must be present.

## Asexual Reproduction

Asexual reproduction occurs in prokaryotic microorganisms (bacteria) and in some eukaryotic single-celled and multi-celled organisms. There are a

number of ways that animals reproduce asexually.

## **Fission**

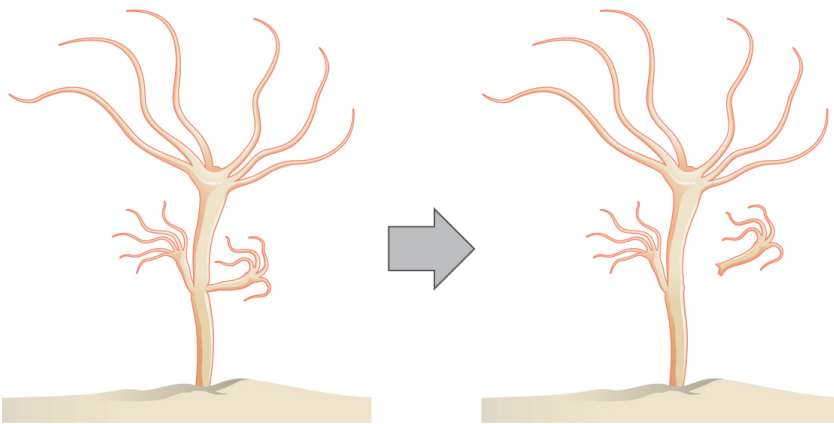
**Fission**, also called binary fission, occurs in prokaryotic microorganisms and in some invertebrate, multi-celled organisms. After a period of growth, an organism splits into two separate organisms. Some unicellular eukaryotic organisms undergo binary fission by mitosis. In other organisms, part of the individual separates and forms a second individual. This process occurs, for example, in many asteroid echinoderms through splitting of the central disk. Some sea anemones and some coral polyps ([link](#)) also reproduce through fission.



Coral polyps reproduce asexually by fission. (credit: G. P. Schmahl, NOAA FGBNMS Manager)

## **Budding**

**Budding** is a form of asexual reproduction that results from the outgrowth of a part of a cell or body region leading to a separation from the original organism into two individuals. Budding occurs commonly in some invertebrate animals such as corals and hydras. In hydras, a bud forms that develops into an adult and breaks away from the main body, as illustrated in [\[link\]](#), whereas in coral budding, the bud does not detach and multiplies as part of a new colony.



Hydra reproduce asexually through budding.

**Note:**

Link to Learning



Watch a video of a hydra budding.

[https://www.openstaxcollege.org/l/budding\\_hydra](https://www.openstaxcollege.org/l/budding_hydra)



## Fragmentation

**Fragmentation** is the breaking of the body into two parts with subsequent regeneration. If the animal is capable of fragmentation, and the part is big enough, a separate individual will regrow.

For example, in many sea stars, asexual reproduction is accomplished by fragmentation. [\[link\]](#) illustrates a sea star for which an arm of the individual is broken off and regenerates a new sea star. Fisheries workers have been known to try to kill the sea stars eating their clam or oyster beds by cutting them in half and throwing them back into the ocean. Unfortunately for the workers, the two parts can each regenerate a new half, resulting in twice as many sea stars to prey upon the oysters and clams. Fragmentation also occurs in annelid worms, turbellarians, and poriferans.



Sea stars can reproduce through fragmentation. The large arm, a fragment from another sea star, is developing into a new individual.

Note that in fragmentation, there is generally a noticeable difference in the size of the individuals, whereas in fission, two individuals of approximate size are formed.

## **Parthenogenesis**

**Parthenogenesis** is a form of asexual reproduction where an egg develops into a complete individual without being fertilized. The resulting offspring can be either haploid or diploid, depending on the process and the species. Parthenogenesis occurs in invertebrates such as water flees, rotifers, aphids, stick insects, some ants, wasps, and bees. Bees use parthenogenesis to produce haploid males (drones). If eggs are fertilized, diploid females develop, and if the fertilized eggs are fed special diet (so called royal jelly), a queen is produced.

Some vertebrate animals—such as certain reptiles, amphibians, and fish—also reproduce through parthenogenesis. Although more common in plants, parthenogenesis has been observed in animal species that were segregated by sex in terrestrial or marine zoos. Two female Komodo dragons, a hammerhead shark, and a blacktip shark have produced parthenogenic young when the females have been isolated from males.

## **Sexual Reproduction**

Sexual reproduction is the combination of (usually haploid) reproductive cells from two individuals to form a third (usually diploid) unique offspring. Sexual reproduction produces offspring with novel combinations of genes. This can be an adaptive advantage in unstable or unpredictable environments. As humans, we are used to thinking of animals as having two separate sexes—male and female—determined at conception. However, in the animal kingdom, there are many variations on this theme.

## **Hermaphroditism**

**Hermaphroditism** occurs in animals where one individual has both male and female reproductive parts. Invertebrates such as earthworms, slugs, tapeworms and snails, shown in [\[link\]](#), are often hermaphroditic. Hermaphrodites may self-fertilize or may mate with another of their species, fertilizing each other and both producing offspring. Self fertilization is common in animals that have limited mobility or are not motile, such as barnacles and clams.



Many snails are hermaphrodites. When two individuals mate, they can produce up to one hundred eggs each. (credit: Assaf Shtilman)

## **Sex Determination**

Mammalian sex determination is determined genetically by the presence of X and Y chromosomes. Individuals homozygous for X (XX) are female and heterozygous individuals (XY) are male. The presence of a Y chromosome causes the development of male characteristics and its absence results in

female characteristics. The XY system is also found in some insects and plants.

Avian sex determination is dependent on the presence of Z and W chromosomes. Homozygous for Z (ZZ) results in a male and heterozygous (ZW) results in a female. The W appears to be essential in determining the sex of the individual, similar to the Y chromosome in mammals. Some fish, crustaceans, insects (such as butterflies and moths), and reptiles use this system.

The sex of some species is not determined by genetics but by some aspect of the environment. Sex determination in some crocodiles and turtles, for example, is often dependent on the temperature during critical periods of egg development. This is referred to as environmental sex determination, or more specifically as temperature-dependent sex determination. In many turtles, cooler temperatures during egg incubation produce males and warm temperatures produce females. In some crocodiles, moderate temperatures produce males and both warm and cool temperatures produce females. In some species, sex is both genetic- and temperature-dependent.

Individuals of some species change their sex during their lives, alternating between male and female. If the individual is female first, it is termed protogyny or “first female,” if it is male first, its termed protandry or “first male.” Oysters, for example, are born male, grow, and become female and lay eggs; some oyster species change sex multiple times.

## **Section Summary**

Reproduction may be asexual when one individual produces genetically identical offspring, or sexual when the genetic material from two individuals is combined to produce genetically diverse offspring. Asexual reproduction occurs through fission, budding, and fragmentation. Sexual reproduction may mean the joining of sperm and eggs within animals' bodies or it may mean the release of sperm and eggs into the environment. An individual may be one sex, or both; it may start out as one sex and switch during its life, or it may stay male or female.

## Review Questions

### Exercise:

#### Problem:

Which form of reproduction is thought to be best in a stable environment?

- a. asexual
- b. sexual
- c. budding
- d. parthenogenesis

---

#### Solution:

A

### Exercise:

#### Problem:

Which form of reproduction can result from damage to the original animal?

- a. asexual
- b. fragmentation
- c. budding
- d. parthenogenesis

---

#### Solution:

B

### Exercise:

**Problem:**

Which form of reproduction is useful to an animal with little mobility that reproduces sexually?

- a. fission
- b. budding
- c. parthenogenesis
- d. hermaphroditism

---

**Solution:**

D

**Exercise:****Problem:**

Genetically unique individuals are produced through \_\_\_\_\_.

- a. sexual reproduction
- b. parthenogenesis
- c. budding
- d. fragmentation

---

**Solution:**

A

**Free Response****Exercise:****Problem:**

Why is sexual reproduction useful if only half the animals can produce offspring and two separate cells must be combined to form a third?

---

**Solution:**

Sexual reproduction produces a new combination of genes in the offspring that may better enable them to survive changes in the environment and assist in the survival of the species.

**Exercise:****Problem:**

What determines which sex will result in offspring of birds and mammals?

---

**Solution:**

The presence of the W chromosome in birds determines femaleness and the presence of the Y chromosome in mammals determines maleness. The absence of those chromosomes and the homogeneity of the offspring (ZZ or XX) leads to the development of the other sex.

## Glossary

asexual reproduction

form of reproduction that produces offspring that are genetically identical to the parent

budding

form of asexual reproduction that results from the outgrowth of a part of a cell leading to a separation from the original animal into two individuals

fission

(also, binary fission) method by which multicellular organisms increase in size or asexual reproduction in which a unicellular organism splits into two separate organisms by mitosis

fragmentation

cutting or fragmenting of the original animal into parts and the growth of a separate animal from each part

hermaphroditism

state of having both male and female reproductive parts within the same individual

parthenogenesis

form of asexual reproduction where an egg develops into a complete individual without being fertilized

sexual reproduction

mixing of genetic material from two individuals to produce genetically unique offspring



## Fertilization

By the end of this section, you will be able to:

- Discuss internal and external methods of fertilization
- Describe the methods used by animals for development of offspring during gestation
- Describe the anatomical adaptations that occurred in animals to facilitate reproduction

Sexual reproduction starts with the combination of a sperm and an egg in a process called fertilization. This can occur either inside (**internal fertilization**) or outside (**external fertilization**) the body of the female. Humans provide an example of the former whereas seahorse reproduction is an example of the latter.

### External Fertilization

External fertilization usually occurs in aquatic environments where both eggs and sperm are released into the water. After the sperm reaches the egg, fertilization takes place. Most external fertilization happens during the process of spawning where one or several females release their eggs and the male(s) release sperm in the same area, at the same time. The release of the reproductive material may be triggered by water temperature or the length of daylight. Nearly all fish spawn, as do crustaceans (such as crabs and shrimp), mollusks (such as oysters), squid, and echinoderms (such as sea urchins and sea cucumbers). [\[link\]](#) shows salmon spawning in a shallow stream. Frogs, like those shown in [\[link\]](#), corals, squid, and octopuses also spawn.



Salmon reproduce through spawning. (credit: Dan Bennett)



During sexual reproduction in toads, the male grasps the female from behind and externally fertilizes the eggs

as they are deposited. (credit:  
"OakleyOriginals"/Flickr)

Pairs of fish that are not broadcast spawners may exhibit courtship behavior. This allows the female to select a particular male. The trigger for egg and sperm release (spawning) causes the egg and sperm to be placed in a small area, enhancing the possibility of fertilization.

External fertilization in an aquatic environment protects the eggs from drying out. Broadcast spawning can result in a greater mixture of the genes within a group, leading to higher genetic diversity and a greater chance of species survival in a hostile environment. For sessile aquatic organisms like sponges, broadcast spawning is the only mechanism for fertilization and colonization of new environments. The presence of the fertilized eggs and developing young in the water provides opportunities for predation resulting in a loss of offspring. Therefore, millions of eggs must be produced by individuals, and the offspring produced through this method must mature rapidly. The survival rate of eggs produced through broadcast spawning is low.

## **Internal Fertilization**

Internal fertilization occurs most often in land-based animals, although some aquatic animals also use this method. There are three ways that offspring are produced following internal fertilization. In **oviparity**, fertilized eggs are laid outside the female's body and develop there, receiving nourishment from the yolk that is a part of the egg. This occurs in most bony fish, many reptiles, some cartilaginous fish, most amphibians, two mammals, and all birds. Reptiles and insects produce leathery eggs, while birds and turtles produce eggs with high concentrations of calcium carbonate in the shell, making them hard. Chicken eggs are an example of this second type.

In **ovoviparity**, fertilized eggs are retained in the female, but the embryo obtains its nourishment from the egg's yolk and the young are fully

developed when they are hatched. This occurs in some bony fish (like the guppy *Lebistes reticulatus*), some sharks, some lizards, some snakes (such as the garter snake *Thamnophis sirtalis*), some vipers, and some invertebrate animals (like the Madagascar hissing cockroach *Gromphadorhina portentosa*).

In **viviparity** the young develop within the female, receiving nourishment from the mother's blood through a placenta. The offspring develops in the female and is born alive. This occurs in most mammals, some cartilaginous fish, and a few reptiles.

Internal fertilization has the advantage of protecting the fertilized egg from dehydration on land. The embryo is isolated within the female, which limits predation on the young. Internal fertilization enhances the fertilization of eggs by a specific male. Fewer offspring are produced through this method, but their survival rate is higher than that for external fertilization.

## **The Evolution of Reproduction**

Once multicellular organisms evolved and developed specialized cells, some also developed tissues and organs with specialized functions. An early development in reproduction occurred in the Annelids. These organisms produce sperm and eggs from undifferentiated cells in their coelom and store them in that cavity. When the coelom becomes filled, the cells are released through an excretory opening or by the body splitting open. Reproductive organs evolved with the development of gonads that produce sperm and eggs. These cells went through meiosis, an adaption of mitosis, which reduced the number of chromosomes in each reproductive cell by half, while increasing the number of cells through cell division.

Complete reproductive systems were developed in insects, with separate sexes. Sperm are made in testes and then travel through coiled tubes to the epididymis for storage. Eggs mature in the ovary. When they are released from the ovary, they travel to the uterine tubes for fertilization. Some insects have a specialized sac, called a **spermatheca**, which stores sperm for later use, sometimes up to a year. Fertilization can be timed with environmental or food conditions that are optimal for offspring survival.

Vertebrates have similar structures, with a few differences. Non-mammals, such as birds and reptiles, have a common body opening, called a **cloaca**, for the digestive, excretory and reproductive systems. Coupling between birds usually involves positioning the cloaca openings opposite each other for transfer of sperm. Mammals have separate openings for the systems in the female and a uterus for support of developing offspring. The uterus has two chambers in species that produce large numbers of offspring at a time, while species that produce one offspring, such as primates, have a single uterus.

Sperm transfer from the male to the female during reproduction ranges from releasing the sperm into the watery environment for external fertilization, to the joining of cloaca in birds, to the development of a penis for direct delivery into the female's vagina in mammals.

## Section Summary

Sexual reproduction starts with the combination of a sperm and an egg in a process called fertilization. This can occur either outside the bodies or inside the female. Both methods have advantages and disadvantages. Once fertilized, the eggs can develop inside the female or outside. If the egg develops outside the body, it usually has a protective covering over it. Animal anatomy evolved various ways to fertilize, hold, or expel the egg. The method of fertilization varies among animals. Some species release the egg and sperm into the environment, some species retain the egg and receive the sperm into the female body and then expel the developing embryo covered with shell, while still other species retain the developing offspring through the gestation period.

## Review Questions

### Exercise:

**Problem:** External fertilization occurs in which type of environment?

- a. aquatic

- b. forested
- c. savanna
- d. steppe

---

**Solution:**

A

**Exercise:**

**Problem:**

Which term applies to egg development within the female with nourishment derived from a yolk?

- a. oviparity
- b. viviparity
- c. ovoviparity
- d. ovovoparity

---

**Solution:**

C

**Exercise:**

**Problem:**

Which term applies to egg development outside the female with nourishment derived from a yolk?

- a. oviparity
- b. viviparity
- c. ovoviparity
- d. ovovoparity

---

**Solution:**

A

## Free Response

### Exercise:

#### Problem:

What are the advantages and disadvantages of external and internal forms of fertilization?

---

#### Solution:

External fertilization can create large numbers of offspring without requiring specialized delivery or reproductive support organs. Offspring develop and mature quickly compared to internally fertilizing species. A disadvantage is that the offspring are out in the environment and predation can account for large loss of offspring. The embryos are susceptible to changes in the environment, which further depletes their numbers. Internally fertilizing species control their environment and protect their offspring from predators but must have specialized organs to complete these tasks and usually produce fewer embryos.

### Exercise:

#### Problem:

Why would paired external fertilization be preferable to group spawning?

---

#### Solution:

Paired external fertilization allows the female to select the male for mating. It also has a greater chance of fertilization taking place, whereas spawning just puts a large number of sperm and eggs together and random interactions result in the fertilization.

## Glossary

### cloaca

common body opening for the digestive, excretory, and reproductive systems found in non-mammals, such as birds

### external fertilization

fertilization of egg by sperm outside animal body, often during spawning

### internal fertilization

fertilization of egg by sperm inside the body of the female

### oviparity

process by which fertilized eggs are laid outside the female's body and develop there, receiving nourishment from the yolk that is a part of the egg

### ovoviviparity

process by which fertilized eggs are retained within the female; the embryo obtains its nourishment from the egg's yolk and the young are fully developed when they are hatched

### spermatheca

specialized sac that stores sperm for later use

### viviparity

process in which the young develop within the female, receiving nourishment from the mother's blood through a placenta



## Human Reproductive Anatomy and Gametogenesis

By the end of this section, you will be able to:

- Describe human male and female reproductive anatomies
- Discuss the human sexual response
- Describe spermatogenesis and oogenesis and discuss their differences and similarities

As animals became more complex, specific organs and organ systems developed to support specific functions for the organism. The reproductive structures that evolved in land animals allow males and females to mate, fertilize internally, and support the growth and development of offspring.

### Human Reproductive Anatomy

The reproductive tissues of male and female humans develop similarly *in utero* until a low level of the hormone testosterone is released from male gonads. Testosterone causes the undeveloped tissues to differentiate into male sexual organs. When testosterone is absent, the tissues develop into female sexual tissues. Primitive gonads become testes or ovaries. Tissues that produce a penis in males produce a clitoris in females. The tissue that will become the scrotum in a male becomes the labia in a female; that is, they are homologous structures.

### Male Reproductive Anatomy

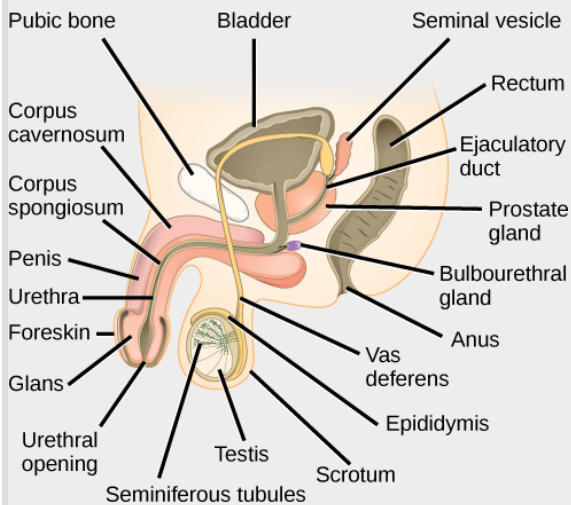
In the male reproductive system, the **scrotum** houses the testicles or testes (singular: testis), including providing passage for blood vessels, nerves, and muscles related to testicular function. The **testes** are a pair of male reproductive organs that produce sperm and some reproductive hormones. Each testis is approximately 2.5 by 3.8 cm (1.5 by 1 in) in size and divided into wedge-shaped lobules by connective tissue called septa. Coiled in each wedge are seminiferous tubules that produce sperm.

Sperm are immobile at body temperature; therefore, the scrotum and penis are external to the body, as illustrated in [\[link\]](#) so that a proper temperature

is maintained for motility. In land mammals, the pair of testes must be suspended outside the body at about  $2^{\circ}\text{C}$  lower than body temperature to produce viable sperm. Infertility can occur in land mammals when the testes do not descend through the abdominal cavity during fetal development.

**Note:**

**Art Connection**



The reproductive structures  
of the human male are  
shown.

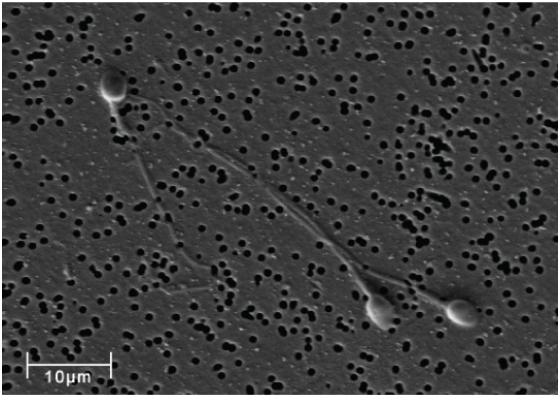
Which of the following statements about the male reproductive system is false?

- a. The vas deferens carries sperm from the testes to the penis.
- b. Sperm mature in seminiferous tubules in the testes.
- c. Both the prostate and the bulbourethral glands produce components of the semen.
- d. The prostate gland is located in the testes.

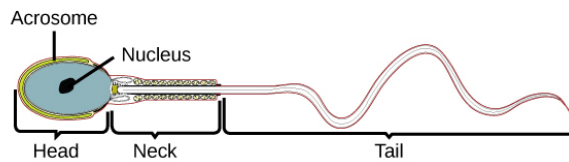
Sperm mature in **seminiferous tubules** that are coiled inside the testes, as illustrated in [\[link\]](#). The walls of the seminiferous tubules are made up of the developing sperm cells, with the least developed sperm at the periphery of the tubule and the fully developed sperm in the lumen. The sperm cells are mixed with “nursemaid” cells called Sertoli cells which protect the germ cells and promote their development. Other cells mixed in the wall of the tubules are the interstitial cells of Leydig. These cells produce high levels of testosterone once the male reaches adolescence.

When the sperm have developed flagella and are nearly mature, they leave the testicles and enter the epididymis, shown in [\[link\]](#). This structure resembles a comma and lies along the top and posterior portion of the testes; it is the site of sperm maturation. The sperm leave the epididymis and enter the vas deferens (or ductus deferens), which carries the sperm, behind the bladder, and forms the ejaculatory duct with the duct from the seminal vesicles. During a vasectomy, a section of the vas deferens is removed, preventing sperm from being passed out of the body during ejaculation and preventing fertilization.

**Semen** is a mixture of sperm and spermatid secretions (about 10 percent of the total) and fluids from accessory glands that contribute most of the semen's volume. Sperm are haploid cells, consisting of a flagellum as a tail, a neck that contains the cell's energy-producing mitochondria, and a head that contains the genetic material. [\[link\]](#) shows a micrograph of human sperm as well as a diagram of the parts of the sperm. An acrosome is found at the top of the head of the sperm. This structure contains lysosomal enzymes that can digest the protective coverings that surround the egg to help the sperm penetrate and fertilize the egg. An ejaculate will contain from two to five milliliters of fluid with from 50–120 million sperm per milliliter.



(a)



(b)

Human sperm, visualized using scanning electron microscopy, have a flagellum, neck, and head. (credit b: modification of work by Mariana Ruiz Villareal; scale-bar data from Matt Russell)

The bulk of the semen comes from the accessory glands associated with the male reproductive system. These are the seminal vesicles, the prostate gland, and the bulbourethral gland, all of which are illustrated in [\[link\]](#). The **seminal vesicles** are a pair of glands that lie along the posterior border of the urinary bladder. The glands make a solution that is thick, yellowish, and alkaline. As sperm are only motile in an alkaline environment, a basic pH is important to reverse the acidity of the vaginal environment. The solution also contains mucus, fructose (a sperm mitochondrial nutrient), a coagulating enzyme, ascorbic acid, and local-acting hormones called prostaglandins. The seminal vesicle glands account for 60 percent of the bulk of semen.

The **penis**, illustrated in [\[link\]](#), is an organ that drains urine from the renal bladder and functions as a copulatory organ during intercourse. The penis contains three tubes of erectile tissue running through the length of the organ. These consist of a pair of tubes on the dorsal side, called the corpus cavernosum, and a single tube of tissue on the ventral side, called the corpus spongiosum. This tissue will become engorged with blood, becoming erect and hard, in preparation for intercourse. The organ is inserted into the vagina culminating with an ejaculation. During intercourse, the smooth muscle sphincters at the opening to the renal bladder close and

prevent urine from entering the penis. An orgasm is a two-stage process: first, glands and accessory organs connected to the testes contract, then semen (containing sperm) is expelled through the urethra during ejaculation. After intercourse, the blood drains from the erectile tissue and the penis becomes flaccid.

The walnut-shaped **prostate gland** surrounds the urethra, the connection to the urinary bladder. It has a series of short ducts that directly connect to the urethra. The gland is a mixture of smooth muscle and glandular tissue. The muscle provides much of the force needed for ejaculation to occur. The glandular tissue makes a thin, milky fluid that contains citrate (a nutrient), enzymes, and prostate specific antigen (PSA). PSA is a proteolytic enzyme that helps to liquefy the ejaculate several minutes after release from the male. Prostate gland secretions account for about 30 percent of the bulk of semen.

The **bulbourethral gland**, or Cowper’s gland, releases its secretion prior to the release of the bulk of the semen. It neutralizes any acid residue in the urethra left over from urine. This usually accounts for a couple of drops of fluid in the total ejaculate and may contain a few sperm. Withdrawal of the penis from the vagina before ejaculation to prevent pregnancy may not work if sperm are present in the bulbourethral gland secretions. The location and functions of the male reproductive organs are summarized in [\[link\]](#).

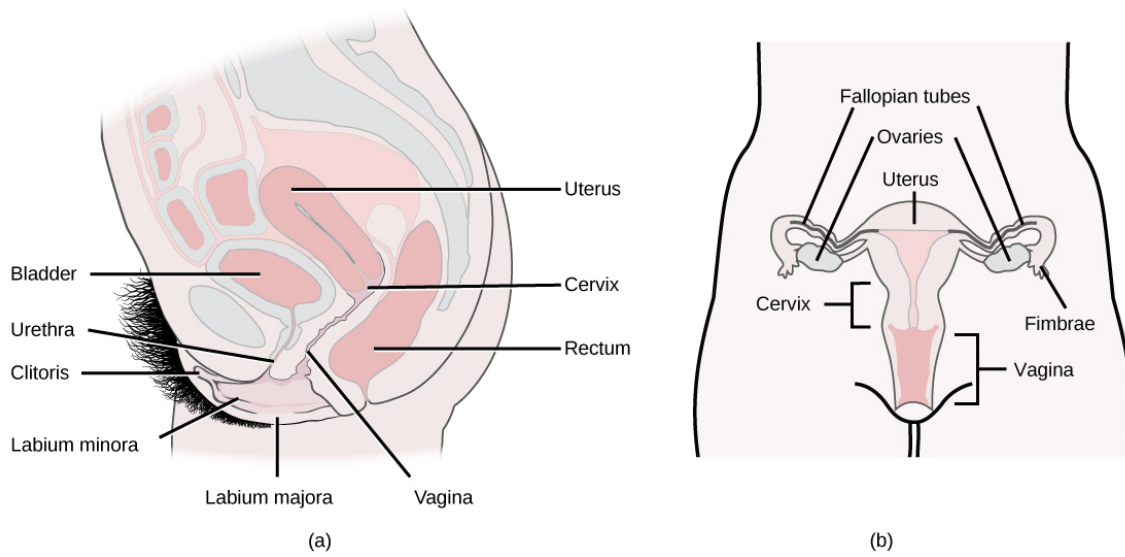
Male Reproductive Anatomy		
Organ	Location	Function
Scrotum	External	Carry and support testes

Male Reproductive Anatomy		
Organ	Location	Function
Penis	External	Deliver urine, copulating organ
Testes	Internal	Produce sperm and male hormones
Seminal Vesicles	Internal	Contribute to semen production
Prostate Gland	Internal	Contribute to semen production
Bulbourethral Glands	Internal	Clean urethra at ejaculation

## Female Reproductive Anatomy

A number of reproductive structures are exterior to the female's body. These include the breasts and the vulva, which consists of the mons pubis, clitoris, labia majora, labia minora, and the vestibular glands, all illustrated in [\[link\]](#). The location and functions of the female reproductive organs are summarized in [\[link\]](#). The vulva is an area associated with the vestibule which includes the structures found in the inguinal (groin) area of women. The mons pubis is a round, fatty area that overlies the pubic symphysis. The **clitoris** is a structure with erectile tissue that contains a large number of sensory nerves and serves as a source of stimulation during intercourse. The **labia majora** are a pair of elongated folds of tissue that run posterior from the mons pubis and enclose the other components of the vulva. The labia majora derive from the same tissue that produces the scrotum in a male. The **labia minora** are thin folds of tissue centrally located within the labia majora. These labia protect the openings to the vagina and urethra. The

mons pubis and the anterior portion of the labia majora become covered with hair during adolescence; the labia minora is hairless. The greater vestibular glands are found at the sides of the vaginal opening and provide lubrication during intercourse.



The reproductive structures of the human female are shown.  
(credit a: modification of work by Gray's Anatomy; credit b:  
modification of work by CDC)

Female Reproductive Anatomy		
Organ	Location	Function
Clitoris	External	Sensory organ

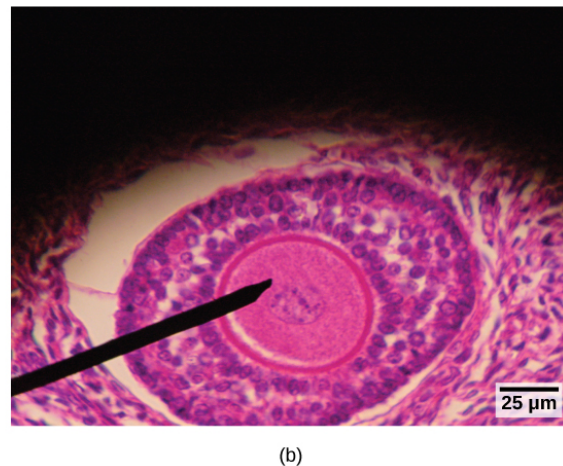
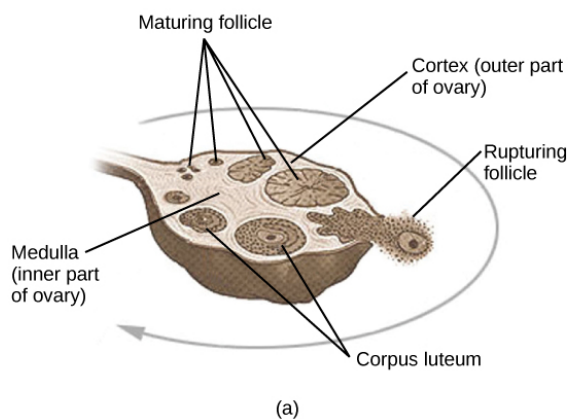
Female Reproductive Anatomy		
Organ	Location	Function
Mons pubis	External	Fatty area overlying pubic bone
Labia majora	External	Covers labia minora
Labia minora	External	Covers vestibule
Greater vestibular glands	External	Secrete mucus; lubricate vagina
Breast	External	Produce and deliver milk
Ovaries	Internal	Carry and develop eggs
Oviducts (Fallopian tubes)	Internal	Transport egg to uterus
Uterus	Internal	Support developing embryo
Vagina	Internal	Common tube for intercourse, birth canal, passing menstrual flow

The breasts consist of mammary glands and fat. The size of the breast is determined by the amount of fat deposited behind the gland. Each gland consists of 15 to 25 lobes that have ducts that empty at the nipple and that supply the nursing child with nutrient- and antibody-rich milk to aid development and protect the child.

Internal female reproductive structures include ovaries, oviducts, the **uterus**, and the vagina, shown in [\[link\]](#). The pair of ovaries is held in place in the abdominal cavity by a system of ligaments. Ovaries consist of a



medulla and cortex: the medulla contains nerves and blood vessels to supply the cortex with nutrients and remove waste. The outer layers of cells of the cortex are the functional parts of the ovaries. The cortex is made up of follicular cells that surround eggs that develop during fetal development *in utero*. During the menstrual period, a batch of follicular cells develops and prepares the eggs for release. At ovulation, one follicle ruptures and one egg is released, as illustrated in [\[link\]](#)a.



Oocytes develop in (a) follicles, located in the ovary. At the beginning of the menstrual cycle, the follicle matures. At ovulation, the follicle ruptures, releasing the egg. The follicle becomes a corpus luteum, which eventually degenerates. The (b) follicle in this light micrograph has an oocyte at its center. (credit a: modification of work by NIH; scale-bar data from Matt Russell)

The **oviducts**, or fallopian tubes, extend from the uterus in the lower abdominal cavity to the ovaries, but they are not in contact with the ovaries. The lateral ends of the oviducts flare out into a trumpet-like structure and have a fringe of finger-like projections called fimbriae, illustrated in [\[link\]](#)b. When an egg is released at ovulation, the fimbriae help the non-motile egg enter into the tube and passage to the uterus. The walls of the oviducts are ciliated and are made up mostly of smooth muscle. The cilia beat toward

the middle, and the smooth muscle contracts in the same direction, moving the egg toward the uterus. Fertilization usually takes place within the oviducts and the developing embryo is moved toward the uterus for development. It usually takes the egg or embryo a week to travel through the oviduct. Sterilization in women is called a tubal ligation; it is analogous to a vasectomy in males in that the oviducts are severed and sealed.

The uterus is a structure about the size of a woman's fist. This is lined with an endometrium rich in blood vessels and mucus glands. The uterus supports the developing embryo and fetus during gestation. The thickest portion of the wall of the uterus is made of smooth muscle. Contractions of the smooth muscle in the uterus aid in passing the baby through the vagina during labor. A portion of the lining of the uterus sloughs off during each menstrual period, and then builds up again in preparation for an implantation. Part of the uterus, called the cervix, protrudes into the top of the vagina. The cervix functions as the birth canal.

The **vagina** is a muscular tube that serves several purposes. It allows menstrual flow to leave the body. It is the receptacle for the penis during intercourse and the vessel for the delivery of offspring. It is lined by stratified squamous epithelial cells to protect the underlying tissue.

## **Sexual Response during Intercourse**

The sexual response in humans is both psychological and physiological. Both sexes experience sexual arousal through psychological and physical stimulation. There are four phases of the sexual response. During phase one, called excitement, vasodilation leads to vasocongestion in erectile tissues in both men and women. The nipples, clitoris, labia, and penis engorge with blood and become enlarged. Vaginal secretions are released to lubricate the vagina to facilitate intercourse. During the second phase, called the plateau, stimulation continues, the outer third of the vaginal wall enlarges with blood, and breathing and heart rate increase.

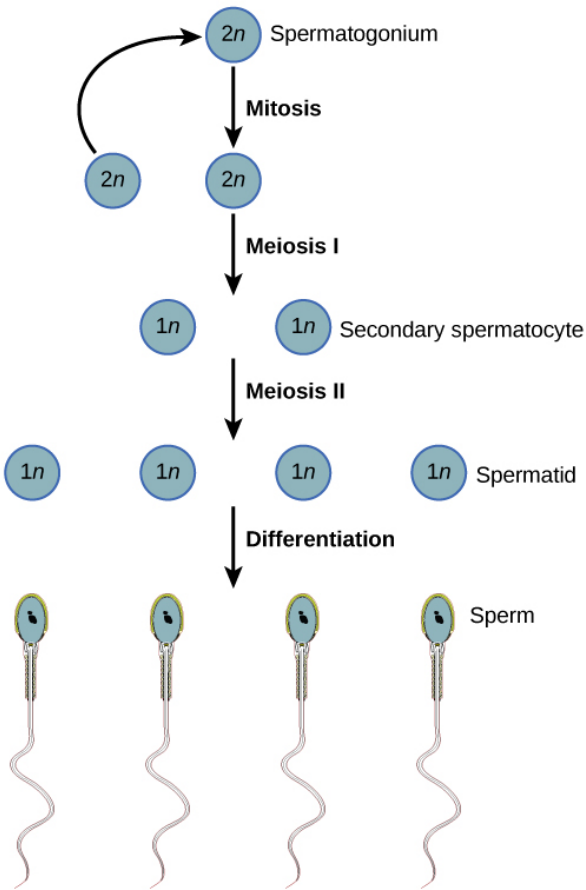
During phase three, or orgasm, rhythmic, involuntary contractions of muscles occur in both sexes. In the male, the reproductive accessory glands

and tubules constrict placing semen in the urethra, then the urethra contracts expelling the semen through the penis. In women, the uterus and vaginal muscles contract in waves that may last slightly less than a second each. During phase four, or resolution, the processes described in the first three phases reverse themselves and return to their normal state. Men experience a refractory period in which they cannot maintain an erection or ejaculate for a period of time ranging from minutes to hours.

## **Gametogenesis (Spermatogenesis and Oogenesis)**

Gametogenesis, the production of sperm and eggs, takes place through the process of meiosis. During meiosis, two cell divisions separate the paired chromosomes in the nucleus and then separate the chromatids that were made during an earlier stage of the cell's life cycle. Meiosis produces haploid cells with half of each pair of chromosomes normally found in diploid cells. The production of sperm is called **spermatogenesis** and the production of eggs is called **oogenesis**.

### **Spermatogenesis**



During spermatogenesis, four sperm result from each primary spermatocyte.

Spermatogenesis, illustrated in [\[link\]](#), occurs in the wall of the seminiferous tubules ([\[link\]](#)), with stem cells at the periphery of the tube and the spermatozoa at the lumen of the tube. Immediately under the capsule of the tubule are diploid, undifferentiated cells. These stem cells, called spermatogonia (singular: spermatagonium), go through mitosis with one offspring going on to differentiate into a sperm cell and the other giving rise to the next generation of sperm.

Meiosis starts with a cell called a primary spermatocyte. At the end of the first meiotic division, a haploid cell is produced called a secondary spermatocyte. This cell is haploid and must go through another meiotic cell

division. The cell produced at the end of meiosis is called a spermatid and when it reaches the lumen of the tubule and grows a flagellum, it is called a sperm cell. Four sperm result from each primary spermatocyte that goes through meiosis.

Stem cells are deposited during gestation and are present at birth through the beginning of adolescence, but in an inactive state. During adolescence, gonadotropic hormones from the anterior pituitary cause the activation of these cells and the production of viable sperm. This continues into old age.

**Note:**

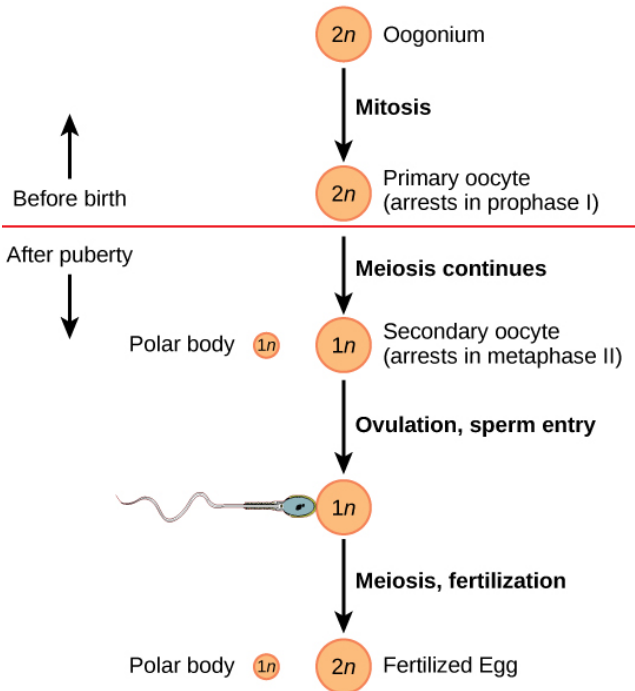
Link to Learning



Visit [this site](#) to see the process of spermatogenesis.

## Oogenesis

Oogenesis, illustrated in [\[link\]](#), occurs in the outermost layers of the ovaries. As with sperm production, oogenesis starts with a germ cell, called an oogonium (plural: oogonia), but this cell undergoes mitosis to increase in number, eventually resulting in up to about one to two million cells in the embryo.



The process of oogenesis occurs in the ovary's outermost layer.

The cell starting meiosis is called a primary oocyte, as shown in [\[link\]](#). This cell will start the first meiotic division and be arrested in its progress in the first prophase stage. At the time of birth, all future eggs are in the prophase stage. At adolescence, anterior pituitary hormones cause the development of a number of follicles in an ovary. This results in the primary oocyte finishing the first meiotic division. The cell divides unequally, with most of the cellular material and organelles going to one cell, called a secondary oocyte, and only one set of chromosomes and a small amount of cytoplasm going to the other cell. This second cell is called a polar body and usually dies. A secondary meiotic arrest occurs, this time at the metaphase II stage. At ovulation, this secondary oocyte will be released and travel toward the uterus through the oviduct. If the secondary oocyte is fertilized, the cell continues through the meiosis II, producing a second polar body and a fertilized egg containing all 46 chromosomes of a human being, half of them coming from the sperm.

Egg production begins before birth, is arrested during meiosis until puberty, and then individual cells continue through at each menstrual cycle. One egg is produced from each meiotic process, with the extra chromosomes and chromatids going into polar bodies that degenerate and are reabsorbed by the body.

## Section Summary

As animals became more complex, specific organs and organ systems developed to support specific functions for the organism. The reproductive structures that evolved in land animals allow males and females to mate, fertilize internally, and support the growth and development of offspring. Processes developed to produce reproductive cells that had exactly half the number of chromosomes of each parent so that new combinations would have the appropriate amount of genetic material. Gametogenesis, the production of sperm (spermatogenesis) and eggs (oogenesis), takes place through the process of meiosis.

### Exercise:

#### Problem:

[\[link\]](#) Which of the following statements about the male reproductive system is false?

- a. The vas deferens carries sperm from the testes to the penis.
- b. Sperm mature in seminiferous tubules in the testes.
- c. Both the prostate and the bulbourethral glands produce components of the semen.
- d. The prostate gland is located in the testes.

---

#### Solution:

[\[link\]](#) D

## Review Questions

**Exercise:**

**Problem:** Sperm are produced in the \_\_\_\_\_.

- a. scrotum
  - b. seminal vesicles
  - c. seminiferous tubules
  - d. prostate gland
- 

**Solution:**

C

**Exercise:**

**Problem:** Most of the bulk of semen is made by the \_\_\_\_\_.

- a. scrotum
  - b. seminal vesicles
  - c. seminiferous tubules
  - d. prostate gland
- 

**Solution:**

C

**Exercise:**

**Problem:** Which of the following cells in spermatogenesis is diploid?

- a. primary spermatocyte
- b. secondary spermatocyte
- c. spermatid
- d. sperm



---

**Solution:**

A

**Exercise:**

**Problem:**

Which female organ has the same embryonic origin as the penis?

- a. clitoris
- b. labia majora
- c. greater vestibular glands
- d. vagina

---

**Solution:**

A

**Exercise:**

**Problem:**

Which female organ has an endometrial lining that will support a developing baby?

- a. labia minora
- b. breast
- c. ovaries
- d. uterus

---

**Solution:**

D

**Exercise:**

**Problem:**

How many eggs are produced as a result of one meiotic series of cell divisions?

- a. one
- b. two
- c. three
- d. four

---

**Solution:**

A

**Free Response****Exercise:**

**Problem:** Describe the phases of the human sexual response.

---

**Solution:**

In phase one (excitement), vasodilation leads to vasocongestion and enlargement of erectile tissues. Vaginal secretions are released to lubricate the vagina during intercourse. In phase two (plateau), stimulation continues, the outer third of the vaginal wall enlarges with blood, and breathing and heart rate increase. In phase three (orgasm), rhythmic, involuntary contractions of muscles occur. In the male, reproductive accessory glands and tubules constrict, depositing semen in the urethra; then, the urethra contracts, expelling the semen through the penis. In women, the uterus and vaginal muscles contract in waves that may last slightly less than a second each. In phase four (resolution), the processes listed in the first three phases reverse themselves and return to their normal state. Men experience a refractory period in which they cannot maintain an erection or

ejaculate for a period of time ranging from minutes to hours. Women do not experience a refractory period.

**Exercise:**

**Problem:**

Compare spermatogenesis and oogenesis as to timing of the processes and the number and type of cells finally produced.

---

**Solution:**

Stem cells are laid down in the male during gestation and lie dormant until adolescence. Stem cells in the female increase to one to two million and enter the first meiotic division and are arrested in prophase. At adolescence, spermatogenesis begins and continues until death, producing the maximum number of sperm with each meiotic division. Oogenesis continues again at adolescence in batches of oogonia with each menstrual cycle. These oogonia finish the first meiotic division, producing a primary oocyte with most of the cytoplasm and its contents, and a second cell called a polar body containing 23 chromosomes. The second meiotic division results in a secondary oocyte and a second oocyte. At ovulation, a mature haploid egg is released. If this egg is fertilized, it finishes the second meiotic division, including the chromosomes donated by the sperm in the finished cell. This is a diploid, fertilized egg.

## **Glossary**

bulbourethral gland

secretion that cleanses the urethra prior to ejaculation

clitoris

sensory structure in females; stimulated during sexual arousal

labia majora

large folds of tissue covering the inguinal area

labia minora

smaller folds of tissue within the labia majora

oogenesis

process of producing haploid eggs

oviduct

(also, fallopian tube) muscular tube connecting the uterus with the ovary area

penis

male reproductive structure for urine elimination and copulation

prostate gland

structure that is a mixture of smooth muscle and glandular material and that contributes to semen

scrotum

sac containing testes; exterior to the body

semen

fluid mixture of sperm and supporting materials

seminal vesicle

secretory accessory gland in males; contributes to semen

seminiferous tubule

site of sperm production in testes

spermatogenesis

process of producing haploid sperm

testes

pair of reproductive organs in males

uterus

environment for developing embryo and fetus

vagina

muscular tube for the passage of menstrual flow, copulation, and birth of offspring

## Hormonal Control of Human Reproduction

By the end of this chapter, you will be able to:

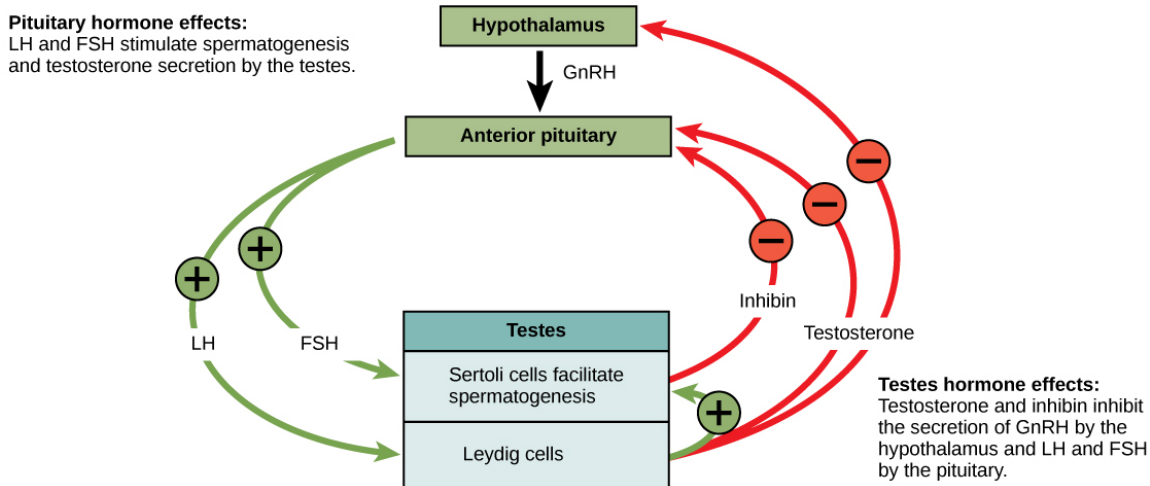
- Describe the roles of male and female reproductive hormones
- Discuss the interplay of the ovarian and menstrual cycles
- Describe the process of menopause

The human male and female reproductive cycles are controlled by the interaction of hormones from the hypothalamus and anterior pituitary with hormones from reproductive tissues and organs. In both sexes, the hypothalamus monitors and causes the release of hormones from the pituitary gland. When the reproductive hormone is required, the hypothalamus sends a **gonadotropin-releasing hormone (GnRH)** to the anterior pituitary. This causes the release of **follicle stimulating hormone (FSH)** and **luteinizing hormone (LH)** from the anterior pituitary into the blood. Note that the body must reach puberty in order for the adrenals to release the hormones that must be present for GnRH to be produced. Although FSH and LH are named after their functions in female reproduction, they are produced in both sexes and play important roles in controlling reproduction. Other hormones have specific functions in the male and female reproductive systems.

### Male Hormones

At the onset of puberty, the hypothalamus causes the release of FSH and LH into the male system for the first time. FSH enters the testes and stimulates the **Sertoli cells** to begin facilitating spermatogenesis using negative feedback, as illustrated in [\[link\]](#). LH also enters the testes and stimulates the **interstitial cells of Leydig** to make and release testosterone into the testes and the blood.

**Testosterone**, the hormone responsible for the secondary sexual characteristics that develop in the male during adolescence, stimulates spermatogenesis. These secondary sex characteristics include a deepening of the voice, the growth of facial, axillary, and pubic hair, and the beginnings of the sex drive.



Hormones control sperm production in a negative feedback system.

A negative feedback system occurs in the male with rising levels of testosterone acting on the hypothalamus and anterior pituitary to inhibit the release of GnRH, FSH, and LH. The Sertoli cells produce the hormone **inhibin**, which is released into the blood when the sperm count is too high. This inhibits the release of GnRH and FSH, which will cause spermatogenesis to slow down. If the sperm count reaches 20 million/ml, the Sertoli cells cease the release of inhibin, and the sperm count increases.

## Female Hormones

The control of reproduction in females is more complex. As with the male, the anterior pituitary hormones cause the release of the hormones FSH and LH. In addition, estrogens and progesterone are released from the developing follicles. **Estrogen** is the reproductive hormone in females that assists in endometrial regrowth, ovulation, and calcium absorption; it is also responsible for the secondary sexual characteristics of females. These include breast development, flaring of the hips, and a shorter period necessary for bone maturation. **Progesterone** assists in endometrial regrowth and inhibition of FSH and LH release.

In females, FSH stimulates development of egg cells, called ova, which develop in structures called follicles. Follicle cells produce the hormone inhibin, which inhibits FSH production. LH also plays a role in the development of ova, induction of ovulation, and stimulation of estradiol and progesterone production by the ovaries. Estradiol and progesterone are steroid hormones that prepare the body for pregnancy. Estradiol produces secondary sex characteristics in females, while both estradiol and progesterone regulate the menstrual cycle.

## **The Ovarian Cycle and the Menstrual Cycle**

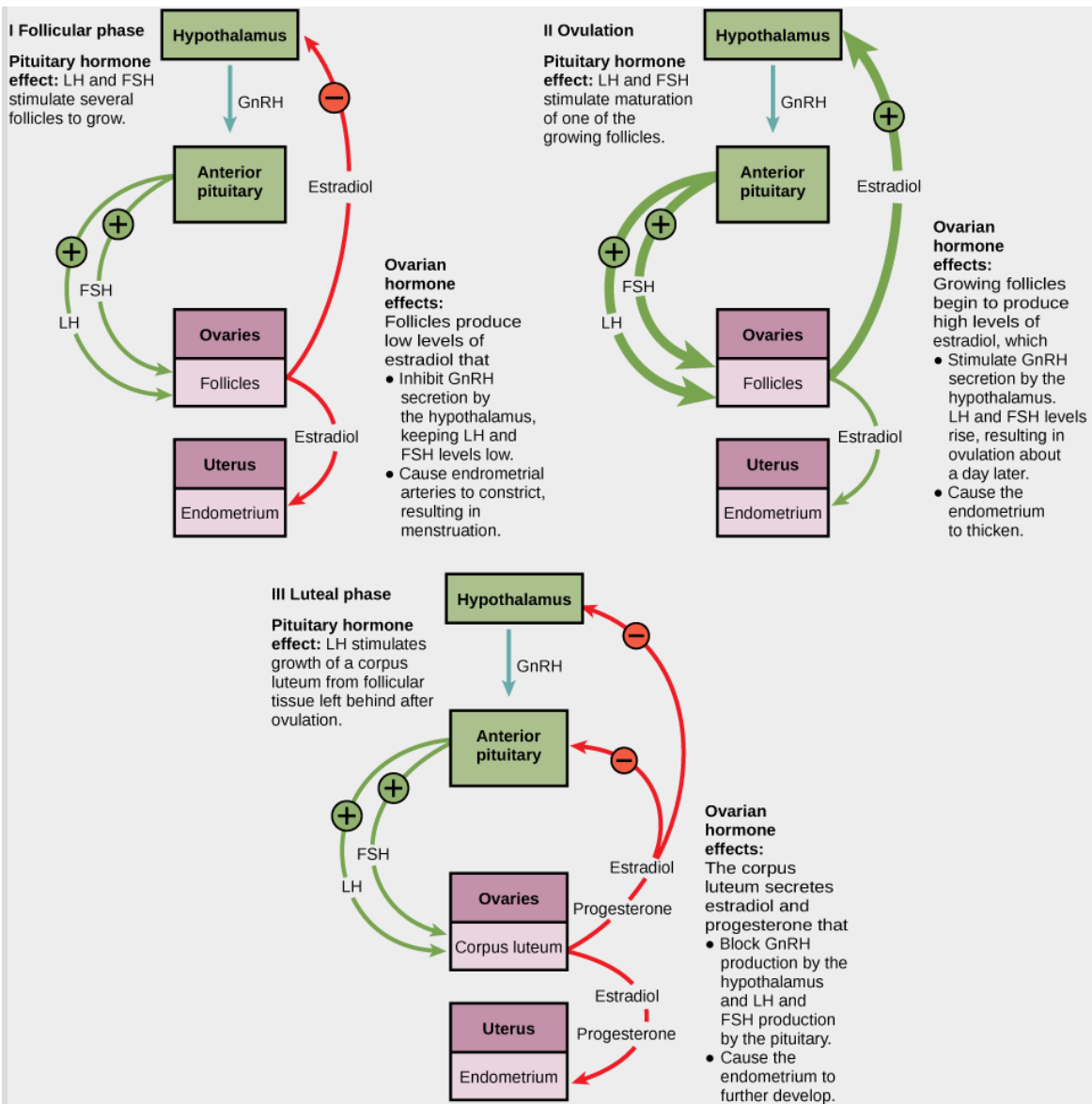
The **ovarian cycle** governs the preparation of endocrine tissues and release of eggs, while the **menstrual cycle** governs the preparation and maintenance of the uterine lining. These cycles occur concurrently and are coordinated over a 22–32 day cycle, with an average length of 28 days.

The first half of the ovarian cycle is the follicular phase shown in [\[link\]](#). Slowly rising levels of FSH and LH cause the growth of follicles on the surface of the ovary. This process prepares the egg for ovulation. As the follicles grow, they begin releasing estrogens and a low level of progesterone. Progesterone maintains the endometrium to help ensure pregnancy. The trip through the fallopian tube takes about seven days. At this stage of development, called the morula, there are 30-60 cells. If pregnancy implantation does not occur, the lining is sloughed off. After about five days, estrogen levels rise and the menstrual cycle enters the proliferative phase. The endometrium begins to regrow, replacing the blood vessels and glands that deteriorated during the end of the last cycle.

### **Note:**

Art Connection





The ovarian and menstrual cycles of female reproduction are regulated by hormones produced by the hypothalamus, pituitary, and ovaries.

Which of the following statements about hormone regulation of the female reproductive cycle is false?

- LH and FSH are produced in the pituitary, and estradiol and progesterone are produced in the ovaries.

- b. Estradiol and progesterone secreted from the corpus luteum cause the endometrium to thicken.
- c. Both progesterone and estradiol are produced by the follicles.
- d. Secretion of GnRH by the hypothalamus is inhibited by low levels of estradiol but stimulated by high levels of estradiol.

Just prior to the middle of the cycle (approximately day 14), the high level of estrogen causes FSH and especially LH to rise rapidly, then fall. The spike in LH causes **ovulation**: the most mature follicle, like that shown in [\[link\]](#), ruptures and releases its egg. The follicles that did not rupture degenerate and their eggs are lost. The level of estrogen decreases when the extra follicles degenerate.



This mature egg follicle may rupture and release an egg. (credit: scale-

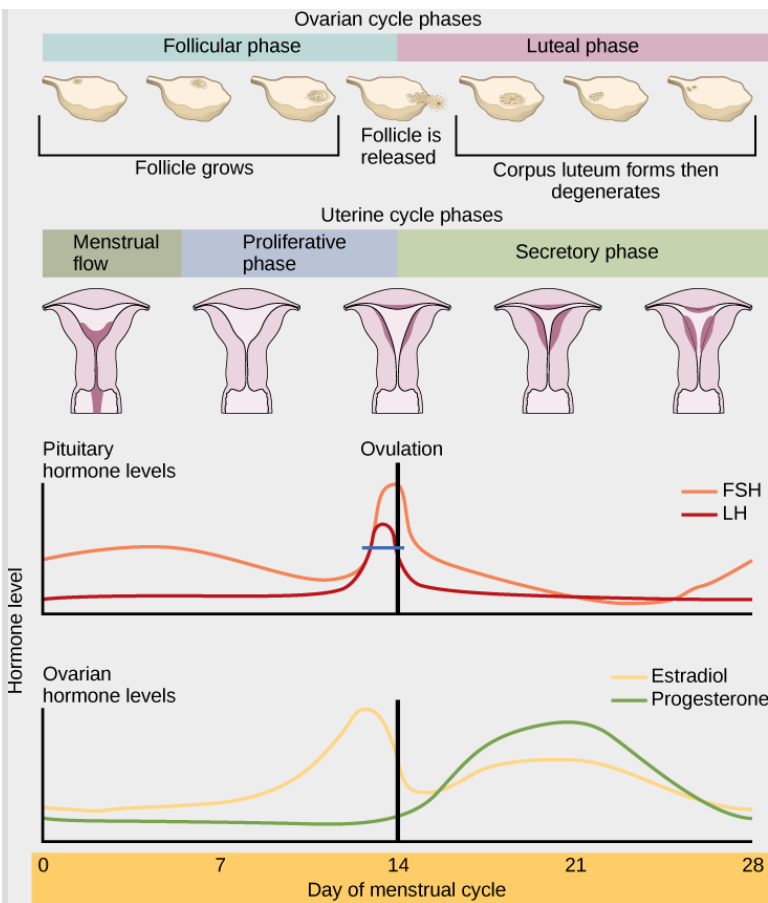
bar data from Matt  
Russell)

Following ovulation, the ovarian cycle enters its luteal phase, illustrated in [\[link\]](#) and the menstrual cycle enters its secretory phase, both of which run from about day 15 to 28. The luteal and secretory phases refer to changes in the ruptured follicle. The cells in the follicle undergo physical changes and produce a structure called a corpus luteum. The corpus luteum produces estrogen and progesterone. The progesterone facilitates the regrowth of the uterine lining and inhibits the release of further FSH and LH. The uterus is being prepared to accept a fertilized egg, should it occur during this cycle. The inhibition of FSH and LH prevents any further eggs and follicles from developing, while the progesterone is elevated. The level of estrogen produced by the corpus luteum increases to a steady level for the next few days.

If no fertilized egg is implanted into the uterus, the corpus luteum degenerates and the levels of estrogen and progesterone decrease. The endometrium begins to degenerate as the progesterone levels drop, initiating the next menstrual cycle. The decrease in progesterone also allows the hypothalamus to send GnRH to the anterior pituitary, releasing FSH and LH and starting the cycles again. [\[link\]](#) visually compares the ovarian and uterine cycles as well as the commensurate hormone levels.

**Note:**

Art Connection



Rising and falling hormone levels result in progression of the ovarian and menstrual cycles. (credit: modification of work by Mikael Häggström)

Which of the following statements about the menstrual cycle is false?

- Progesterone levels rise during the luteal phase of the ovarian cycle and the secretory phase of the uterine cycle.
- Menstruation occurs just after LH and FSH levels peak.
- Menstruation occurs after progesterone levels drop.
- Estrogen levels rise before ovulation, while progesterone levels rise after.

## Menopause

As women approach their mid-40s to mid-50s, their ovaries begin to lose their sensitivity to FSH and LH. Menstrual periods become less frequent and finally cease; this is **menopause**. There are still eggs and potential follicles on the ovaries, but without the stimulation of FSH and LH, they will not produce a viable egg to be released. The outcome of this is the inability to have children.

The side effects of menopause include hot flashes, heavy sweating (especially at night), headaches, some hair loss, muscle pain, vaginal dryness, insomnia, depression, weight gain, and mood swings. Estrogen is involved in calcium metabolism and, without it, blood levels of calcium decrease. To replenish the blood, calcium is lost from bone which may decrease the bone density and lead to osteoporosis. Supplementation of estrogen in the form of hormone replacement therapy (HRT) can prevent bone loss, but the therapy can have negative side effects. While HRT is thought to give some protection from colon cancer, osteoporosis, heart disease, macular degeneration, and possibly depression, its negative side effects include increased risk of: stroke or heart attack, blood clots, breast cancer, ovarian cancer, endometrial cancer, gall bladder disease, and possibly dementia.

### **Note:**

#### Career Connection

#### **Reproductive Endocrinologist**

A reproductive endocrinologist is a physician who treats a variety of hormonal disorders related to reproduction and infertility in both men and women. The disorders include menstrual problems, infertility, pregnancy loss, sexual dysfunction, and menopause. Doctors may use fertility drugs, surgery, or assisted reproductive techniques (ART) in their therapy. ART involves the use of procedures to manipulate the egg or sperm to facilitate reproduction, such as *in vitro* fertilization.

Reproductive endocrinologists undergo extensive medical training, first in a four-year residency in obstetrics and gynecology, then in a three-year

fellowship in reproductive endocrinology. To be board certified in this area, the physician must pass written and oral exams in both areas.

## Section Summary

The male and female reproductive cycles are controlled by hormones released from the hypothalamus and anterior pituitary as well as hormones from reproductive tissues and organs. The hypothalamus monitors the need for the FSH and LH hormones made and released from the anterior pituitary. FSH and LH affect reproductive structures to cause the formation of sperm and the preparation of eggs for release and possible fertilization. In the male, FSH and LH stimulate Sertoli cells and interstitial cells of Leydig in the testes to facilitate sperm production. The Leydig cells produce testosterone, which also is responsible for the secondary sexual characteristics of males. In females, FSH and LH cause estrogen and progesterone to be produced. They regulate the female reproductive system which is divided into the ovarian cycle and the menstrual cycle. Menopause occurs when the ovaries lose their sensitivity to FSH and LH and the female reproductive cycles slow to a stop.

## Art Connections

### Exercise:

#### Problem:

[\[link\]](#) Which of the following statements about hormone regulation of the female reproductive cycle is false?

- a. LH and FSH are produced in the pituitary, and estradiol and progesterone are produced in the ovaries.
- b. Estradiol and progesterone secreted from the corpus luteum cause the endometrium to thicken.
- c. Both progesterone and estradiol are produced by the follicles.
- d. Secretion of GnRH by the hypothalamus is inhibited by low levels of estradiol but stimulated by high levels of estradiol.

---

**Solution:**

[\[link\]](#) C

**Exercise:**

**Problem:**

[\[link\]](#) Which of the following statements about the menstrual cycle is false?

- a. Progesterone levels rise during the luteal phase of the ovarian cycle and the secretory phase of the uterine cycle.
- b. Menstruation occurs just after LH and FSH levels peak.
- c. Menstruation occurs after progesterone levels drop.
- d. Estrogen levels rise before ovulation, while progesterone levels rise after.

---

**Solution:**

[\[link\]](#) B

## Review Questions

**Exercise:**

**Problem:** Which hormone causes Leydig cells to make testosterone?

- a. FSH
- b. LH
- c. inhibin
- d. estrogen

---

**Solution:**

A

**Exercise:**

**Problem:** Which hormone causes FSH and LH to be released?

- a. testosterone
  - b. estrogen
  - c. GnRH
  - d. progesterone
- 

**Solution:**

C

**Exercise:**

**Problem:** Which hormone signals ovulation?

- a. FSH
  - b. LH
  - c. inhibin
  - d. estrogen
- 

**Solution:**

B

**Exercise:**

**Problem:**

Which hormone causes the re-growth of the endometrial lining of the uterus?

- a. testosterone
- b. estrogen
- c. GnRH



d. progesterone

---

**Solution:**

D

## Free Response

**Exercise:**

**Problem:**

If male reproductive pathways are not cyclical, how are they controlled?

---

**Solution:**

Negative feedback in the male system is supplied through two hormones: inhibin and testosterone. Inhibin is produced by Sertoli cells when the sperm count exceeds set limits. The hormone inhibits GnRH and FSH, decreasing the activity of the Sertoli cells. Increased levels of testosterone affect the release of both GnRH and LH, decreasing the activity of the Leydig cells, resulting in decreased testosterone and sperm production.

**Exercise:**

**Problem:**

Describe the events in the ovarian cycle leading up to ovulation.

---

**Solution:**

Low levels of progesterone allow the hypothalamus to send GnRH to the anterior pituitary and cause the release of FSH and LH. FSH stimulates follicles on the ovary to grow and prepare the eggs for ovulation. As the follicles increase in size, they begin to release estrogen and a low level of progesterone into the blood. The level of

estrogen rises to a peak, causing a spike in the concentration of LH. This causes the most mature follicle to rupture and ovulation occurs.

## **Glossary**

### **estrogen**

reproductive hormone in females that assists in endometrial regrowth, ovulation, and calcium absorption

### **follicle stimulating hormone (FSH)**

reproductive hormone that causes sperm production in men and follicle development in women

### **gonadotropin-releasing hormone (GnRH)**

hormone from the hypothalamus that causes the release of FSH and LH from the anterior pituitary

### **inhibin**

hormone made by Sertoli cells; provides negative feedback to hypothalamus in control of FSH and GnRH release

### **interstitial cell of Leydig**

cell in seminiferous tubules that makes testosterone

### **luteinizing hormone (LH)**

reproductive hormone in both men and women, causes testosterone production in men and ovulation and lactation in women

### **menopause**

loss of reproductive capacity in women due to decreased sensitivity of the ovaries to FSH and LH

### **menstrual cycle**

cycle of the degradation and re-growth of the endometrium

### **ovarian cycle**

cycle of preparation of egg for ovulation and the conversion of the follicle to the corpus luteum

ovulation

release of the egg by the most mature follicle

progesterone

reproductive hormone in women; assists in endometrial re-growth and inhibition of FSH and LH release

Sertoli cell

cell in seminiferous tubules that assists developing sperm and makes inhibin

testosterone

reproductive hormone in men that assists in sperm production and promoting secondary sexual characteristics

## Human Pregnancy and Birth

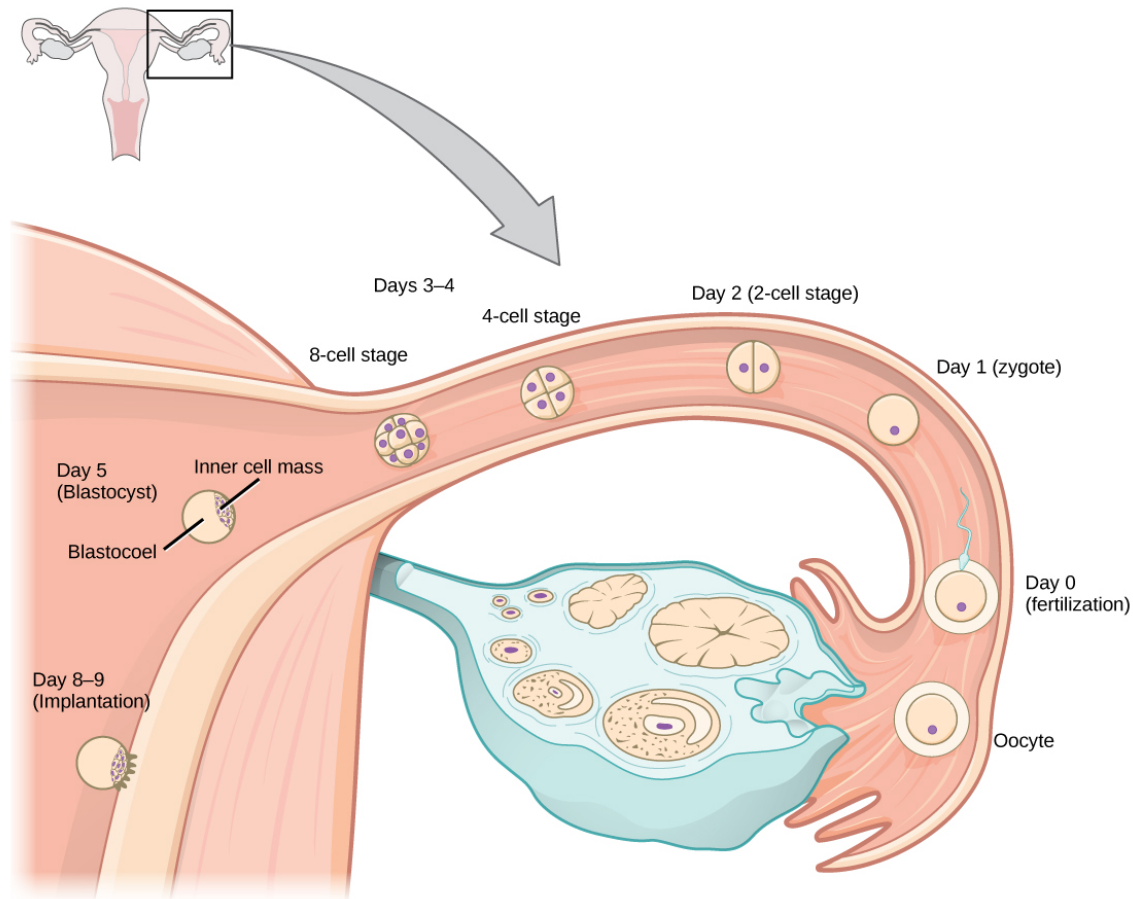
By the end of this section, you will be able to:

- Explain fetal development during the three trimesters of gestation
- Describe labor and delivery
- Compare the efficacy and duration of various types of contraception
- Discuss causes of infertility and the therapeutic options available

Pregnancy begins with the fertilization of an egg and continues through to the birth of the individual. The length of time of **gestation** varies among animals, but is very similar among the great apes: human gestation is 266 days, while chimpanzee gestation is 237 days, a gorilla's is 257 days, and orangutan gestation is 260 days long. The fox has a 57-day gestation. Dogs and cats have similar gestations averaging 60 days. The longest gestation for a land mammal is an African elephant at 640 days. The longest gestations among marine mammals are the beluga and sperm whales at 460 days.

## Human Gestation

Twenty-four hours before fertilization, the egg has finished meiosis and becomes a mature oocyte. When fertilized (at conception) the egg becomes known as a zygote. The zygote travels through the oviduct to the uterus ([link](#)). The developing embryo must implant into the wall of the uterus within seven days, or it will deteriorate and die. The outer layers of the zygote (blastocyst) grow into the endometrium by digesting the endometrial cells, and wound healing of the endometrium closes up the blastocyst into the tissue. Another layer of the blastocyst, the chorion, begins releasing a hormone called **human beta chorionic gonadotropin ( $\beta$ -HCG)** which makes its way to the corpus luteum and keeps that structure active. This ensures adequate levels of progesterone that will maintain the endometrium of the uterus for the support of the developing embryo. Pregnancy tests determine the level of  $\beta$ -HCG in urine or serum. If the hormone is present, the test is positive.



In humans, fertilization occurs soon after the oocyte leaves the ovary. Implantation occurs eight or nine days later.(credit: Ed Uthman)

The gestation period is divided into three equal periods or trimesters. During the first two to four weeks of the first trimester, nutrition and waste are handled by the endometrial lining through diffusion. As the trimester progresses, the outer layer of the embryo begins to merge with the endometrium, and the **placenta** forms. This organ takes over the nutrient and waste requirements of the embryo and fetus, with the mother's blood passing nutrients to the placenta and removing waste from it. Chemicals from the fetus, such as bilirubin, are processed by the mother's liver for elimination. Some of the mother's immunoglobulins will pass through the placenta, providing passive immunity against some potential infections.

Internal organs and body structures begin to develop during the first trimester. By five weeks, limb buds, eyes, the heart, and liver have been basically formed. By eight weeks, the term fetus applies, and the body is essentially formed, as shown in [\[link\]](#). The individual is about five centimeters (two inches) in length and many of the organs, such as the lungs and liver, are not yet functioning. Exposure to any toxins is especially dangerous during the first trimester, as all of the body's organs and structures are going through initial development. Anything that affects that development can have a severe effect on the fetus' survival.



Fetal development is shown at nine weeks gestation. (credit: Ed Uthman)

During the second trimester, the fetus grows to about 30 cm (12 inches), as shown in [\[link\]](#). It becomes active and the mother usually feels the first

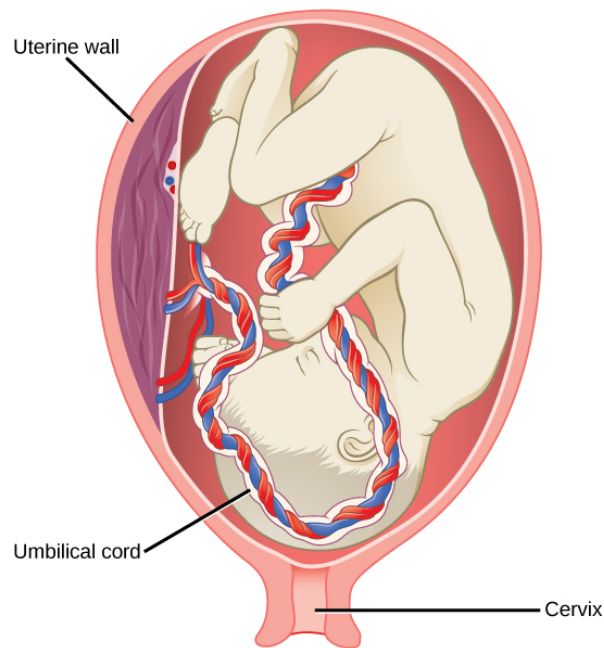
movements. All organs and structures continue to develop. The placenta has taken over the functions of nutrition and waste and the production of estrogen and progesterone from the corpus luteum, which has degenerated. The placenta will continue functioning up through the delivery of the baby.



This fetus is just entering the second trimester, when the placenta takes over more of the functions performed as the baby develops. (credit: National Museum of Health and Medicine)

During the third trimester, the fetus grows to 3 to 4 kg (6 ½ -8 ½ lbs.) and about 50 cm (19-20 inches) long, as illustrated in [\[link\]](#). This is the period of the most rapid growth during the pregnancy. Organ development continues to birth (and some systems, such as the nervous system and liver,

continue to develop after birth). The mother will be at her most uncomfortable during this trimester. She may urinate frequently due to pressure on the bladder from the fetus. There may also be intestinal blockage and circulatory problems, especially in her legs. Clots may form in her legs due to pressure from the fetus on returning veins as they enter the abdominal cavity.



There is rapid fetal growth during the third trimester.  
(credit: modification of work by Gray's Anatomy)

**Note:**  
Link to Learning





Visit [this site](#) to see the stages of human fetal development.

## Labor and Birth

Labor is the physical efforts of expulsion of the fetus and the placenta from the uterus during birth (parturition). Toward the end of the third trimester, estrogen causes receptors on the uterine wall to develop and bind the hormone oxytocin. At this time, the baby reorients, facing forward and down with the back or crown of the head engaging the cervix (uterine opening). This causes the cervix to stretch and nerve impulses are sent to the hypothalamus, which signals for the release of oxytocin from the posterior pituitary. The oxytocin causes the smooth muscle in the uterine wall to contract. At the same time, the placenta releases prostaglandins into the uterus, increasing the contractions. A positive feedback relay occurs between the uterus, hypothalamus, and the posterior pituitary to assure an adequate supply of oxytocin. As more smooth muscle cells are recruited, the contractions increase in intensity and force.

There are three stages to labor. During stage one, the cervix thins and dilates. This is necessary for the baby and placenta to be expelled during birth. The cervix will eventually dilate to about 10 cm. During stage two, the baby is expelled from the uterus. The uterus contracts and the mother pushes as she compresses her abdominal muscles to aid the delivery. The last stage is the passage of the placenta after the baby has been born and the organ has completely disengaged from the uterine wall. If labor should stop before stage two is reached, synthetic oxytocin, known as Pitocin, can be administered to restart and maintain labor.

An alternative to labor and delivery is the surgical delivery of the baby through a procedure called a Caesarian section. This is major abdominal surgery and can lead to post-surgical complications for the mother, but in some cases it may be the only way to safely deliver the baby.

The mother’s mammary glands go through changes during the third trimester to prepare for lactation and breastfeeding. When the baby begins suckling at the breast, signals are sent to the hypothalamus causing the release of prolactin from the anterior pituitary. Prolactin causes the mammary glands to produce milk. Oxytocin is also released, promoting the release of the milk. The milk contains nutrients for the baby’s development and growth as well as immunoglobulins to protect the child from bacterial and viral infections.

**Contraception and Birth Control**

The prevention of a pregnancy comes under the terms contraception or birth control. Strictly speaking, **contraception** refers to preventing the sperm and egg from joining. Both terms are, however, frequently used interchangeably.

Contraceptive Methods		
Method	Examples	Failure Rate in Typical Use Over 12 Months
Barrier	male condom, female condom, sponge, cervical cap, diaphragm, spermicides	15 to 24%
Hormonal	oral, patch, vaginal ring	8%

Contraceptive Methods		
Method	Examples	Failure Rate in Typical Use Over 12 Months
	injection	3%
	implant	less than 1%
Other	natural family planning	12 to 25%
	withdrawal	27%
	sterilization	less than 1%

[\[link\]](#) lists common methods of contraception. The failure rates listed are not the ideal rates that could be realized, but the typical rates that occur. A failure rate is the number of pregnancies resulting from the method's use over a twelve-month period. Barrier methods, such as condoms, cervical caps, and diaphragms, block sperm from entering the uterus, preventing fertilization. Spermicides are chemicals that are placed in the vagina that kill sperm. Sponges, which are saturated with spermicides, are placed in the vagina at the cervical opening. Combinations of spermicidal chemicals and barrier methods achieve lower failure rates than do the methods when used separately.

Nearly a quarter of the couples using barrier methods, natural family planning, or withdrawal can expect a failure of the method. Natural family planning is based on the monitoring of the menstrual cycle and having intercourse only during times when the egg is not available. A woman's body temperature may rise a degree Celsius at ovulation and the cervical mucus may increase in volume and become more pliable. These changes give a general indication of when intercourse is more or less likely to result in fertilization. Withdrawal involves the removal of the penis from the vagina during intercourse, before ejaculation occurs. This is a risky method

with a high failure rate due to the possible presence of sperm in the bulbourethral gland's secretion, which may enter the vagina prior to removing the penis.

Hormonal methods use synthetic progesterone (sometimes in combination with estrogen), to inhibit the hypothalamus from releasing FSH or LH, and thus prevent an egg from being available for fertilization. The method of administering the hormone affects failure rate. The most reliable method, with a failure rate of less than 1 percent, is the implantation of the hormone under the skin. The same rate can be achieved through the sterilization procedures of vasectomy in the man or of tubal ligation in the woman, or by using an intrauterine device (IUD). IUDs are inserted into the uterus and establish an inflammatory condition that prevents fertilized eggs from implanting into the uterine wall.

Compliance with the contraceptive method is a strong contributor to the success or failure rate of any particular method. The only method that is completely effective at preventing conception is abstinence. The choice of contraceptive method depends on the goals of the woman or couple. Tubal ligation and vasectomy are considered permanent prevention, while other methods are reversible and provide short-term contraception.

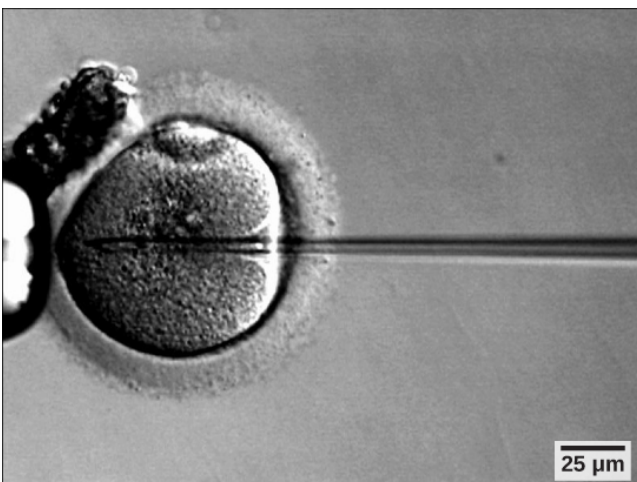
Termination of an existing pregnancy can be spontaneous or voluntary. Spontaneous termination is a miscarriage and usually occurs very early in the pregnancy, usually within the first few weeks. This occurs when the fetus cannot develop properly and the gestation is naturally terminated. Voluntary termination of a pregnancy is an abortion. Laws regulating abortion vary between states and tend to view fetal viability as the criteria for allowing or preventing the procedure.

## **Infertility**

**Infertility** is the inability to conceive a child or carry a child to birth. About 75 percent of causes of infertility can be identified; these include diseases, such as sexually transmitted diseases that can cause scarring of the reproductive tubes in either men or women, or developmental problems frequently related to abnormal hormone levels in one of the individuals.

Inadequate nutrition, especially starvation, can delay menstruation. Stress can also lead to infertility. Short-term stress can affect hormone levels, while long-term stress can delay puberty and cause less frequent menstrual cycles. Other factors that affect fertility include toxins (such as cadmium), tobacco smoking, marijuana use, gonadal injuries, and aging.

If infertility is identified, several assisted reproductive technologies (ART) are available to aid conception. A common type of ART is *in vitro* fertilization (IVF) where an egg and sperm are combined outside the body and then placed in the uterus. Eggs are obtained from the woman after extensive hormonal treatments that prepare mature eggs for fertilization and prepare the uterus for implantation of the fertilized egg. Sperm are obtained from the man and they are combined with the eggs and supported through several cell divisions to ensure viability of the zygotes. When the embryos have reached the eight-cell stage, one or more is implanted into the woman's uterus. If fertilization is not accomplished by simple IVF, a procedure that injects the sperm into an egg can be used. This is called intracytoplasmic sperm injection (ICSI) and is shown in [\[link\]](#). IVF procedures produce a surplus of fertilized eggs and embryos that can be frozen and stored for future use. The procedures can also result in multiple births.



A sperm is inserted into an egg for fertilization during

intracytoplasmic sperm injection  
(ICSI). (credit: scale-bar data from  
Matt Russell)

## Section Summary

Human pregnancy begins with fertilization of an egg and proceeds through the three trimesters of gestation. The labor process has three stages (contractions, delivery of the fetus, expulsion of the placenta), each propelled by hormones. The first trimester lays down the basic structures of the body, including the limb buds, heart, eyes, and the liver. The second trimester continues the development of all of the organs and systems. The third trimester exhibits the greatest growth of the fetus and culminates in labor and delivery. Prevention of a pregnancy can be accomplished through a variety of methods including barriers, hormones, or other means. Assisted reproductive technologies may help individuals who have infertility problems.

## Review Questions

### Exercise:

#### Problem:

Nutrient and waste requirements for the developing fetus are handled during the first few weeks by:

- a. the placenta
- b. diffusion through the endometrium
- c. the chorion
- d. the blastocyst

---

### Solution:

B

**Exercise:**

**Problem:** Progesterone is made during the third trimester by the:

- a. placenta
  - b. endometrial lining
  - c. chorion
  - d. corpus luteum
- 

**Solution:**

A

**Exercise:**

**Problem:**

Which contraceptive method is 100 percent effective at preventing pregnancy?

- a. condom
  - b. oral hormonal methods
  - c. sterilization
  - d. abstinence
- 

**Solution:**

D

**Exercise:**

**Problem:**

Which type of short term contraceptive method is generally more effective than others?

- a. barrier

- b. hormonal
- c. natural family planning
- d. withdrawal

---

**Solution:**

B

**Exercise:**

**Problem:**

Which hormone is primarily responsible for the contractions during labor?

- a. oxytocin
- b. estrogen
- c.  $\beta$ -HCG
- d. progesterone

---

**Solution:**

A

**Exercise:**

**Problem:**

Major organs begin to develop during which part of human gestation?

- a. fertilization
- b. first trimester
- c. second trimester
- d. third trimester

---

**Solution:**

B



## Free Response

### Exercise:

#### Problem:

Describe the major developments during each trimester of human gestation.

---

#### Solution:

The first trimester lays down the basic structures of the body, including the limb buds, heart, eyes, and the liver. The second trimester continues the development of all of the organs and systems established during the first trimester. The placenta takes over the production of estrogen and high levels of progesterone and handles the nutrient and waste requirements of the fetus. The third trimester exhibits the greatest growth of the fetus, culminating in labor and delivery.

### Exercise:

**Problem:** Describe the stages of labor.

---

#### Solution:

Stage one of labor results in the thinning of the cervix and the dilation of the cervical opening. Stage two delivers the baby, and stage three delivers the placenta.

## Glossary

contraception

(also, birth control) various means used to prevent pregnancy

gestation

length of time for fetal development to birth

human beta chorionic gonadotropin ( $\beta$ -HCG)

hormone produced by the chorion of the zygote that helps to maintain the corpus luteum and elevated levels of progesterone

infertility

inability to conceive, carry, and deliver children

morning sickness

condition in the mother during the first trimester; includes feelings of nausea

placenta

organ that supports the diffusion of nutrients and waste between the mother's and fetus' blood

## Fertilization and Early Embryonic Development

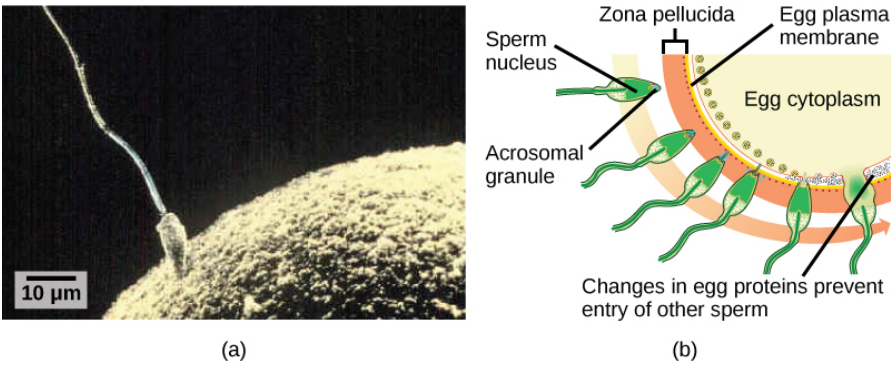
By the end of this section, you will be able to:

- Discuss how fertilization occurs
- Explain how the embryo forms from the zygote
- Discuss the role of cleavage and gastrulation in animal development

The process in which an organism develops from a single-celled zygote to a multi-cellular organism is complex and well-regulated. The early stages of embryonic development are also crucial for ensuring the fitness of the organism.

### Fertilization

Fertilization, pictured in [\[link\]](#)**a** is the process in which gametes (an egg and sperm) fuse to form a zygote. The egg and sperm each contain one set of chromosomes. To ensure that the offspring has only one complete diploid set of chromosomes, only one sperm must fuse with one egg. In mammals, the egg is protected by a layer of extracellular matrix consisting mainly of glycoproteins called the **zona pellucida**. When a sperm binds to the zona pellucida, a series of biochemical events, called the **acrosomal reactions**, take place. In placental mammals, the acrosome contains digestive enzymes that initiate the degradation of the glycoprotein matrix protecting the egg and allowing the sperm plasma membrane to fuse with the egg plasma membrane, as illustrated in [\[link\]](#)**b**. The fusion of these two membranes creates an opening through which the sperm nucleus is transferred into the ovum. The nuclear membranes of the egg and sperm break down and the two haploid genomes condense to form a diploid genome.



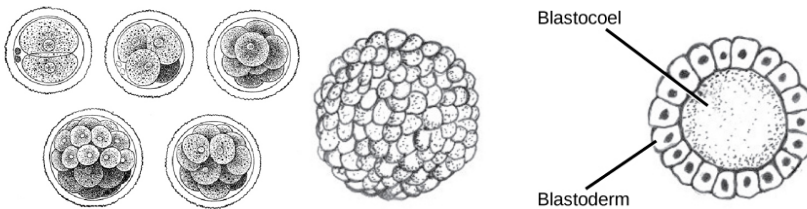
(a) Fertilization is the process in which sperm and egg fuse to form a zygote. (b) Acrosomal reactions help the sperm degrade the glycoprotein matrix protecting the egg and allow the sperm to transfer its nucleus. (credit: (b) modification of work by Mariana Ruiz Villareal; scale-bar data from Matt Russell)

To ensure that no more than one sperm fertilizes the egg, once the acrosomal reactions take place at one location of the egg membrane, the egg releases proteins in other locations to prevent other sperm from fusing with the egg. If this mechanism fails, multiple sperm can fuse with the egg, resulting in **polyspermy**. The resulting embryo is not genetically viable and dies within a few days.

## Cleavage and Blastula Stage

The development of multi-cellular organisms begins from a single-celled zygote, which undergoes rapid cell division to form the blastula. The rapid, multiple rounds of cell division are termed cleavage. Cleavage is illustrated in ([link]a). After the cleavage has produced over 100 cells, the embryo is called a blastula. The blastula is usually a spherical layer of cells (the blastoderm) surrounding a fluid-filled or yolk-filled cavity (the blastocoel). Mammals at this stage form a structure called the blastocyst, characterized by an inner cell mass that is distinct from the surrounding blastula, shown in [link]b. During cleavage, the cells divide without an increase in mass;

that is, one large single-celled zygote divides into multiple smaller cells. Each cell within the blastula is called a blastomere.

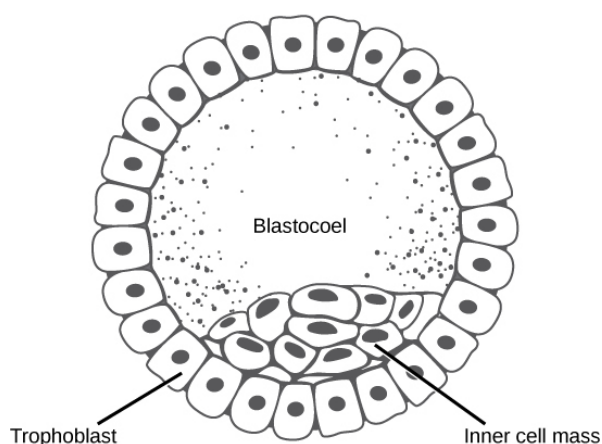


(a) During cleavage, the zygote rapidly divides into multiple cells without increasing in size. (b) The cells rearrange themselves to form a hollow ball with a fluid-filled or yolk-filled cavity called the blastula. (credit a: modification of work by Gray's Anatomy; credit b: modification of work by Pearson Scott Foresman, donated to the Wikimedia Foundation)

Cleavage can take place in two ways: **holoblastic** (total) cleavage or **meroblastic** (partial) cleavage. The type of cleavage depends on the amount of yolk in the eggs. In placental mammals (including humans) where nourishment is provided by the mother's body, the eggs have a very small amount of yolk and undergo holoblastic cleavage. Other species, such as birds, with a lot of yolk in the egg to nourish the embryo during development, undergo meroblastic cleavage.

In mammals, the blastula forms the **blastocyst** in the next stage of development. Here the cells in the blastula arrange themselves in two layers: the **inner cell mass**, and an outer layer called the **trophoblast**. The inner cell mass is also known as the embryoblast and this mass of cells will go on to form the embryo. At this stage of development, illustrated in [\[link\]](#) the inner cell mass consists of embryonic stem cells that will differentiate

into the different cell types needed by the organism. The trophoblast will contribute to the placenta and nourish the embryo.



The rearrangement of the cells in the mammalian blastula to two layers—the inner cell mass and the trophoblast—results in the formation of the blastocyst.

**Note:**

Link to Learning

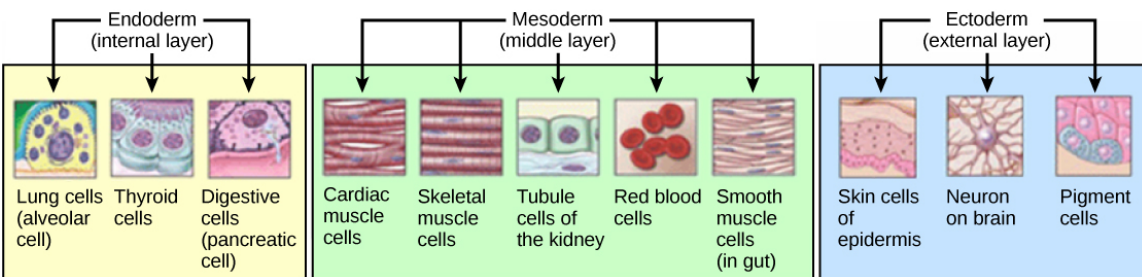


Visit the [Virtual Human Embryo project](#) at the Endowment for Human Development site to step through an interactive that shows the stages of embryo development, including micrographs and rotating 3-D images.

## Gastrulation

The typical blastula is a ball of cells. The next stage in embryonic development is the formation of the body plan. The cells in the blastula rearrange themselves spatially to form three layers of cells. This process is called **gastrulation**. During gastrulation, the blastula folds upon itself to form the three layers of cells. Each of these layers is called a germ layer and each germ layer differentiates into different organ systems.

The three germ layers, shown in [\[link\]](#), are the endoderm, the ectoderm, and the mesoderm. The ectoderm gives rise to the nervous system and the epidermis. The mesoderm gives rise to the muscle cells and connective tissue in the body. The endoderm gives rise to columnar cells found in the digestive system and many internal organs.

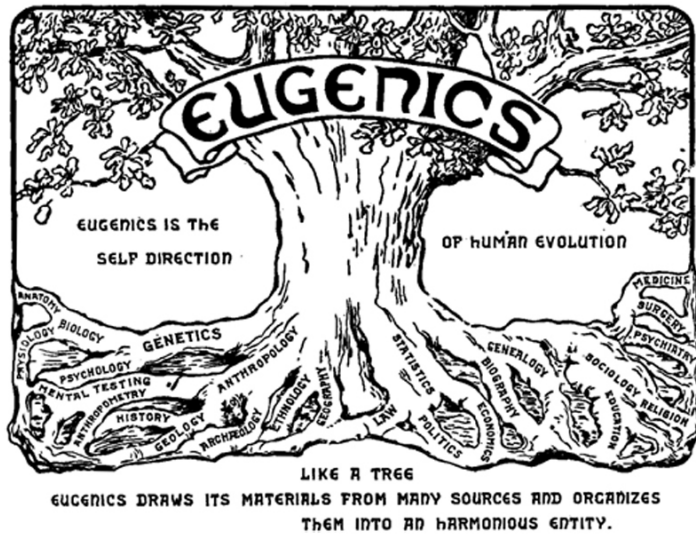


The three germ layers give rise to different cell types in the animal body. (credit: modification of work by NIH, NCBI)

### Note:

Everyday Connection

**Are Designer Babies in Our Future?**



This logo from the Second International Eugenics Conference in New York City in September of 1921 shows how eugenics attempted to merge several fields of study with the goal of producing a genetically superior human race.

If you could prevent your child from getting a devastating genetic disease, would you do it? Would you select the sex of your child or select for their attractiveness, strength, or intelligence? How far would you go to maximize the possibility of resistance to disease? The genetic engineering of a human child, the production of "designer babies" with desirable phenotypic characteristics, was once a topic restricted to science fiction. This is the case no longer: science fiction is now overlapping into science fact. Many phenotypic choices for offspring are already available, with many more likely to be possible in the not too distant future. Which traits should be selected and how they should be selected are topics of much debate within the worldwide medical community. The ethical and moral line is not always clear or agreed upon, and some fear that modern reproductive technologies could lead to a new form of eugenics. Eugenics is the use of information and technology from a variety of sources to improve the genetic makeup of the human race. The goal of



creating genetically superior humans was quite prevalent (although controversial) in several countries during the early 20<sup>th</sup> century, but fell into disrepute when Nazi Germany developed an extensive eugenics program in the 1930's and 40's. As part of their program, the Nazis forcibly sterilized hundreds of thousands of the so-called "unfit" and killed tens of thousands of institutionally disabled people as part of a systematic program to develop a genetically superior race of Germans known as Aryans. Ever since, eugenic ideas have not been as publicly expressed, but there are still those who promote them.

Efforts have been made in the past to control traits in human children using donated sperm from men with desired traits. In fact, eugenicist Robert Klark Graham established a sperm bank in 1980 that included samples exclusively from donors with high IQs. The "genius" sperm bank failed to capture the public's imagination and the operation closed in 1999.

In more recent times, the procedure known as prenatal genetic diagnosis (PGD) has been developed. PGD involves the screening of human embryos as part of the process of *in vitro* fertilization, during which embryos are conceived and grown outside the mother's body for some period of time before they are implanted. The term PGD usually refers to both the diagnosis, selection, and the implantation of the selected embryos.

In the least controversial use of PGD, embryos are tested for the presence of alleles which cause genetic diseases such as sickle cell disease, muscular dystrophy, and hemophilia, in which a single disease-causing allele or pair of alleles has been identified. By excluding embryos containing these alleles from implantation into the mother, the disease is prevented, and the unused embryos are either donated to science or discarded. There are relatively few in the worldwide medical community that question the ethics of this type of procedure, which allows individuals scared to have children because of the alleles they carry to do so successfully. The major limitation to this procedure is its expense. Not usually covered by medical insurance and thus out of reach financially for most couples, only a very small percentage of all live births use such complicated methodologies. Yet, even in cases like these where the ethical issues may seem to be clear-cut, not everyone agrees with the morality of these types of procedures. For example, to those who take the position that human life begins at conception, the discarding of unused embryos, a necessary result of PGD, is unacceptable under any circumstances.

A murkier ethical situation is found in the selection of a child's sex, which is easily performed by PGD. Currently, countries such as Great Britain have banned the selection of a child's sex for reasons other than preventing sex-linked diseases. Other countries allow the procedure for "family balancing", based on the desire of some parents to have at least one child of each sex. Still others, including the United States, have taken a scattershot approach to regulating these practices, essentially leaving it to the individual practicing physician to decide which practices are acceptable and which are not.

Even murkier are rare instances of disabled parents, such as those with deafness or dwarfism, who select embryos via PGD to ensure that they share their disability. These parents usually cite many positive aspects of their disabilities and associated culture as reasons for their choice, which they see as their moral right. To others, to purposely cause a disability in a child violates the basic medical principle of *Primum non nocere*, "first, do no harm." This procedure, although not illegal in most countries, demonstrates the complexity of ethical issues associated with choosing genetic traits in offspring.

Where could this process lead? Will this technology become more affordable and how should it be used? With the ability of technology to progress rapidly and unpredictably, a lack of definitive guidelines for the use of reproductive technologies before they arise might make it difficult for legislators to keep pace once they are in fact realized, assuming the process needs any government regulation at all. Other bioethicists argue that we should only deal with technologies that exist now, and not in some uncertain future. They argue that these types of procedures will always be expensive and rare, so the fears of eugenics and "master" races are unfounded and overstated. The debate continues.

## Section Summary

The early stages of embryonic development begin with fertilization. The process of fertilization is tightly controlled to ensure that only one sperm fuses with one egg. After fertilization, the zygote undergoes cleavage to form the blastula. The blastula, which in some species is a hollow ball of

cells, undergoes a process called gastrulation, in which the three germ layers form. The ectoderm gives rise to the nervous system and the epidermal skin cells, the mesoderm gives rise to the muscle cells and connective tissue in the body, and the endoderm gives rise to columnar cells and internal organs.

## Review Questions

### Exercise:

**Problem:** Which of the following is false?

- a. The endoderm, mesoderm, ectoderm are germ layers.
- b. The trophoblast is a germ layer.
- c. The inner cell mass is a source of embryonic stem cells.
- d. The blastula is often a hollow ball of cells.

---

**Solution:**

B

### Exercise:

**Problem:** During cleavage, the mass of cells:

- a. increases
- b. decreases
- c. doubles with every cell division
- d. does not change significantly

---

**Solution:**

D

## Free Response

**Exercise:****Problem:**

What do you think would happen if multiple sperm fused with one egg?

---

**Solution:**

Multiple sperm can fuse with the egg, resulting in polyspermy. The resulting embryo is not genetically viable and dies within a few days.

**Exercise:****Problem:**

Why do mammalian eggs have a small concentration of yolk, while bird and reptile eggs have a large concentration of yolk?

---

**Solution:**

Mammalian eggs do not need a lot of yolk because the developing fetus obtains nutrients from the mother. Other species, in which the fetus develops outside of the mother's body, such as occurs with birds, require a lot of yolk in the egg to nourish the embryo during development.

**Glossary**

acrosomal reaction

series of biochemical reactions that the sperm uses to break through the zona pellucida

blastocyst

structure formed when cells in the mammalian blastula separate into an inner and outer layer

gastrulation

process in which the blastula folds over itself to form the three germ layers

holoblastic

complete cleavage; takes place in cells with a small amount of yolk

inner cell mass

inner layer of cells in the blastocyst

meroblastic

partial cleavage; takes place in cells with a large amount of yolk

polyspermy

condition in which one egg is fertilized by multiple sperm

trophoblast

outer layer of cells in the blastocyst

zona pellucida

protective layer of glycoproteins on the mammalian egg

## Organogenesis and Vertebrate Formation

By the end of this section, you will be able to:

- Describe the process of organogenesis
- Identify the anatomical axes formed in vertebrates

Gastrulation leads to the formation of the three germ layers that give rise, during further development, to the different organs in the animal body. This process is called **organogenesis**. Organogenesis is characterized by rapid and precise movements of the cells within the embryo.

### Organogenesis

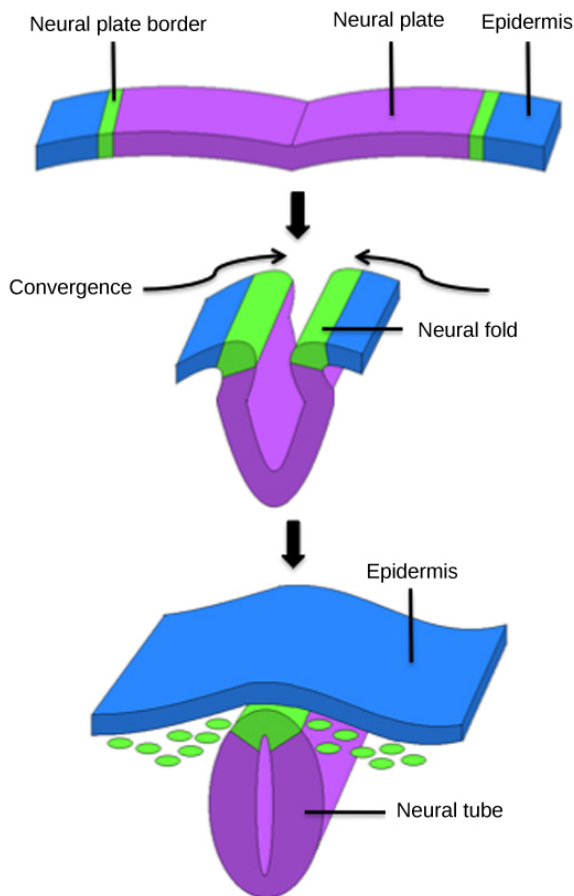
Organs form from the germ layers through the process of differentiation. During differentiation, the embryonic stem cells express specific sets of genes which will determine their ultimate cell type. For example, some cells in the ectoderm will express the genes specific to skin cells. As a result, these cells will differentiate into epidermal cells. The process of differentiation is regulated by cellular signaling cascades.

Scientists study organogenesis extensively in the lab in fruit flies (*Drosophila*) and the nematode *Caenorhabditis elegans*. *Drosophila* have segments along their bodies, and the patterning associated with the segment formation has allowed scientists to study which genes play important roles in organogenesis along the length of the embryo at different time points. The nematode *C.elegans* has roughly 1000 somatic cells and scientists have studied the fate of each of these cells during their development in the nematode life cycle. There is little variation in patterns of cell lineage between individuals, unlike in mammals where cell development from the embryo is dependent on cellular cues.

In vertebrates, one of the primary steps during organogenesis is the formation of the neural system. The ectoderm forms epithelial cells and tissues, and neuronal tissues. During the formation of the neural system, special signaling molecules called growth factors signal some cells at the edge of the ectoderm to become epidermis cells. The remaining cells in the

center form the neural plate. If the signaling by growth factors were disrupted, then the entire ectoderm would differentiate into neural tissue.

The neural plate undergoes a series of cell movements where it rolls up and forms a tube called the **neural tube**, as illustrated in [\[link\]](#). In further development, the neural tube will give rise to the brain and the spinal cord.



The central region of the ectoderm forms the neural tube, which gives rise to the brain and the spinal cord.

The mesoderm that lies on either side of the vertebrate neural tube will develop into the various connective tissues of the animal body. A spatial pattern of gene expression reorganizes the mesoderm into groups of cells called **somites** with spaces between them. The somites, illustrated in [\[link\]](#) will further develop into the ribs, lungs, and segmental (spine) muscle. The mesoderm also forms a structure called the notochord, which is rod-shaped and forms the central axis of the animal body.



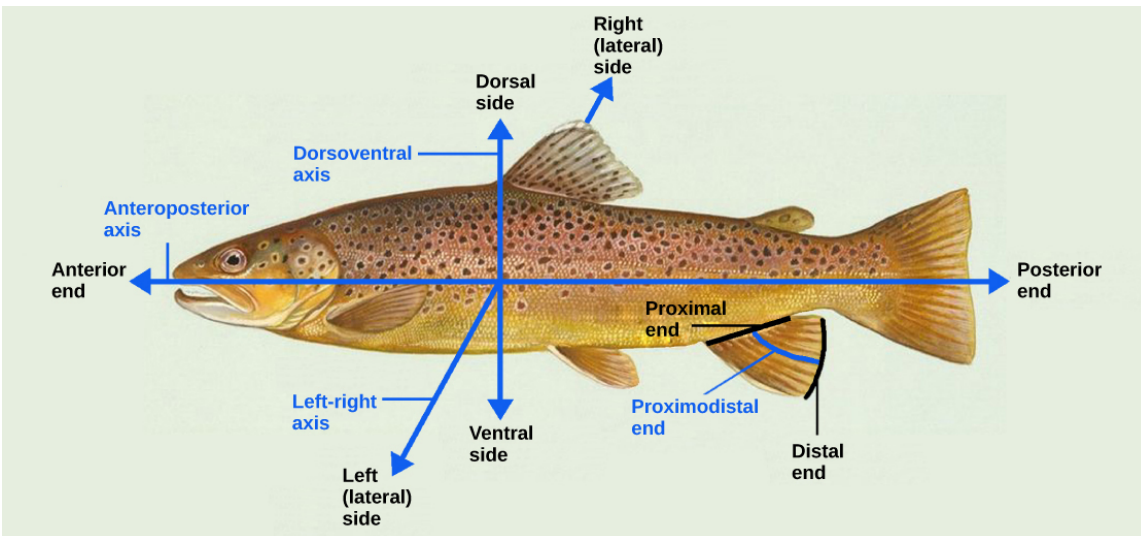
In this five-week old human embryo, somites are segments along the length of the body.  
(credit: modification of work by Ed Uthman)

## Vertebrate Axis Formation

Even as the germ layers form, the ball of cells still retains its spherical shape. However, animal bodies have lateral-medial (left-right), dorsal-



ventral (back-belly), and anterior-posterior (head-feet) axes, illustrated in [\[link\]](#).



Animal bodies have three axes for symmetry. (credit: modification of work by NOAA)

How are these established? In one of the most seminal experiments ever to be carried out in developmental biology, Spemann and Mangold took dorsal cells from one embryo and transplanted them into the belly region of another embryo. They found that the transplanted embryo now had two notochords: one at the dorsal site from the original cells and another at the transplanted site. This suggested that the dorsal cells were genetically programmed to form the notochord and define the axis. Since then, researchers have identified many genes that are responsible for axis formation. Mutations in these genes leads to the loss of symmetry required for organism development.

Animal bodies have externally visible symmetry. However, the internal organs are not symmetric. For example, the heart is on the left side and the liver on the right. The formation of the central left-right axis is an important process during development. This internal asymmetry is established very

early during development and involves many genes. Research is still ongoing to fully understand the developmental implications of these genes.

## Section Summary

Organogenesis is the formation of organs from the germ layers. Each germ layer gives rise to specific tissue types. The first stage is the formation of the neural system in the ectoderm. The mesoderm gives rise to somites and the notochord. Formation of vertebrate axis is another important developmental stage.

## Review Questions

### Exercise:

**Problem:** Which of the following gives rise to the skin cells?

- a. ectoderm
- b. endoderm
- c. mesoderm
- d. none of the above

---

### Solution:

A

### Exercise:

**Problem:** The ribs form from the \_\_\_\_\_.

- a. notochord
- b. neural plate
- c. neural tube
- d. somites

---

**Solution:**

D

**Free Response****Exercise:****Problem:**

Explain how the different germ layers give rise to different tissue types.

---

**Solution:**

Organs form from the germ layers through the process of differentiation. During differentiation, the embryonic stem cells express a specific set of genes that will determine their ultimate fate as a cell type. For example, some cells in the ectoderm will express the genes specific to skin cells. As a result, these cells will differentiate into epidermal cells. The process of differentiation is regulated by cellular signaling cascades.

**Exercise:**

**Problem:** Explain the role of axis formation in development.

---

**Solution:**

Animal bodies have lateral-medial (left-right), dorsal-ventral (back-belly), and anterior-posterior (head-feet) axes. The dorsal cells are genetically programmed to form the notochord and define the axis. There are many genes responsible for axis formation. Mutations in these genes lead to the loss of symmetry required for organism development.

## **Glossary**

neural tube

tube-like structure that forms from the ectoderm and gives rise to the brain and spinal cord

organogenesis

process of organ formation

somite

group of cells separated by small spaces that form from the mesoderm and give rise to connective tissue

## Introduction

class="introduction"

All  
organisms  
are products  
of evolution  
adapted to  
their  
environment.

(a) Saguaro  
(*Carnegiea  
gigantea*)  
can soak up  
750 liters of  
water in a  
single rain  
storm,  
enabling  
these cacti to  
survive the  
dry  
conditions of  
the Sonora  
desert in  
Mexico and  
the  
Southwester  
n United  
States. (b)  
The Andean  
semiaquatic  
lizard  
(*Potamites  
montanicola*)  
discovered in  
Peru in 2010

lives  
between  
1,570 to  
2,100 meters  
in elevation,  
and, unlike  
most lizards,  
is nocturnal  
and swims.

Scientists  
still do not  
know how  
these cold-  
blood  
animals are  
able to move  
in the cold  
(10 to 15°C)  
temperatures  
of the  
Andean  
night. (credit

a:  
modification  
of work by  
Gentry  
George, U.S.

Fish and  
Wildlife  
Service;  
credit b:  
modification  
of work by  
Germán  
Chávez and  
Diego

Vásquez,  
*ZooKeys*)



All species of living organisms, from bacteria to baboons to blueberries, evolved at some point from a different species. Although it may seem that living things today stay much the same, that is not the case—evolution is an ongoing process.

The theory of evolution is the unifying theory of biology, meaning it is the framework within which biologists ask questions about the living world. Its power is that it provides direction for predictions about living things that are borne out in experiment after experiment. The Ukrainian-born American geneticist Theodosius Dobzhansky famously wrote that “nothing makes sense in biology except in the light of evolution.”<sup>[[footnote](#)]</sup> He meant that the tenet that all life has evolved and diversified from a common ancestor is the foundation from which we approach all questions in biology. Theodosius Dobzhansky. “Biology, Molecular and Organismic.” *American Zoologist* 4, no. 4 (1964): 449.

## Understanding Evolution

By the end of this section, you will be able to:

- Describe how the present-day theory of evolution was developed
- Define adaptation
- Explain convergent and divergent evolution
- Describe homologous and vestigial structures
- Discuss misconceptions about the theory of evolution

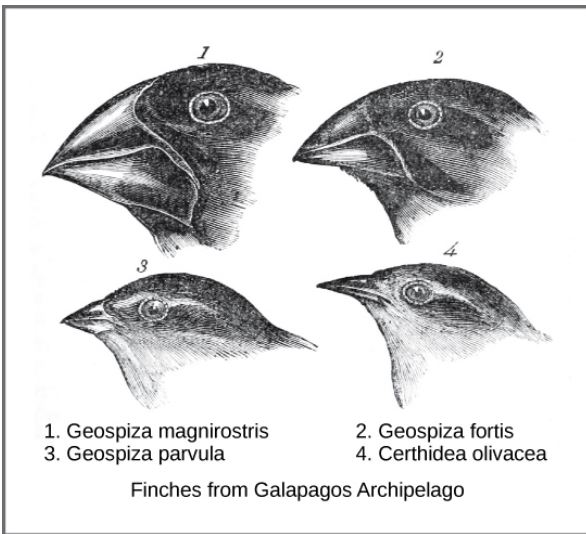
Evolution by natural selection describes a mechanism for how species change over time. That species change had been suggested and debated well before Darwin began to explore this idea. The view that species were static and unchanging was grounded in the writings of Plato, yet there were also ancient Greeks who expressed evolutionary ideas. In the eighteenth century, ideas about the evolution of animals were reintroduced by the naturalist Georges-Louis Leclerc Comte de Buffon who observed that various geographic regions have different plant and animal populations, even when the environments are similar. It was also accepted that there were extinct species.

During this time, James Hutton, a Scottish naturalist, proposed that geological change occurred gradually by the accumulation of small changes from processes operating like they are today over long periods of time. This contrasted with the predominant view that the geology of the planet was a consequence of catastrophic events occurring during a relatively brief past. Hutton's view was popularized in the nineteenth century by the geologist Charles Lyell who became a friend to Darwin. Lyell's ideas were influential on Darwin's thinking: Lyell's notion of the greater age of Earth gave more time for gradual change in species, and the process of change provided an analogy for gradual change in species. In the early nineteenth century, Jean-Baptiste Lamarck published a book that detailed a mechanism for evolutionary change. This mechanism is now referred to as an inheritance of acquired characteristics by which modifications in an individual are caused by its environment, or the use or disuse of a structure during its lifetime, could be inherited by its offspring and thus bring about change in a species. While this mechanism for evolutionary change was discredited, Lamarck's ideas were an important influence on evolutionary thought.



## Charles Darwin and Natural Selection

In the mid-nineteenth century, the actual mechanism for evolution was independently conceived of and described by two naturalists: Charles Darwin and Alfred Russel Wallace. Importantly, each naturalist spent time exploring the natural world on expeditions to the tropics. From 1831 to 1836, Darwin traveled around the world on *H.M.S. Beagle*, including stops in South America, Australia, and the southern tip of Africa. Wallace traveled to Brazil to collect insects in the Amazon rainforest from 1848 to 1852 and to the Malay Archipelago from 1854 to 1862. Darwin's journey, like Wallace's later journeys to the Malay Archipelago, included stops at several island chains, the last being the Galápagos Islands west of Ecuador. On these islands, Darwin observed species of organisms on different islands that were clearly similar, yet had distinct differences. For example, the ground finches inhabiting the Galápagos Islands comprised several species with a unique beak shape ([\[link\]](#)). The species on the islands had a graded series of beak sizes and shapes with very small differences between the most similar. He observed that these finches closely resembled another finch species on the mainland of South America. Darwin imagined that the island species might be species modified from one of the original mainland species. Upon further study, he realized that the varied beaks of each finch helped the birds acquire a specific type of food. For example, seed-eating finches had stronger, thicker beaks for breaking seeds, and insect-eating finches had spear-like beaks for stabbing their prey.



Darwin observed that beak shape varies among finch species. He postulated that the beak of an ancestral species had adapted over time to equip the finches to acquire different food sources.

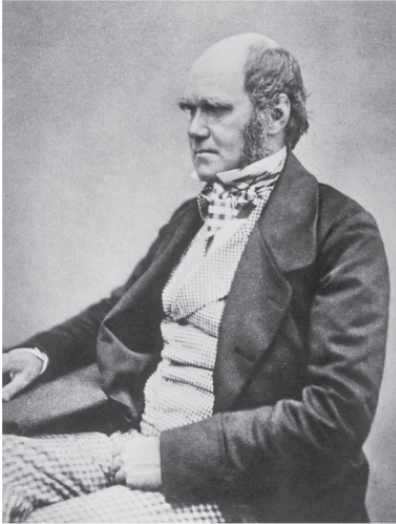
Wallace and Darwin both observed similar patterns in other organisms and they independently developed the same explanation for how and why such changes could take place. Darwin called this mechanism natural selection. **Natural selection**, also known as “survival of the fittest,” is the more prolific reproduction of individuals with favorable traits that survive environmental change because of those traits; this leads to evolutionary change.

For example, a population of giant tortoises found in the Galapagos Archipelago was observed by Darwin to have longer necks than those that lived on other islands with dry lowlands. These tortoises were “selected” because they could reach more leaves and access more food than those with short necks. In times of drought when fewer leaves would be available, those that could reach more leaves had a better chance to eat and survive than those that couldn’t reach the food source. Consequently, long-necked

tortoises would be more likely to be reproductively successful and pass the long-necked trait to their offspring. Over time, only long-necked tortoises would be present in the population.

Natural selection, Darwin argued, was an inevitable outcome of three principles that operated in nature. First, most characteristics of organisms are inherited, or passed from parent to offspring. Although no one, including Darwin and Wallace, knew how this happened at the time, it was a common understanding. Second, more offspring are produced than are able to survive, so resources for survival and reproduction are limited. The capacity for reproduction in all organisms outstrips the availability of resources to support their numbers. Thus, there is competition for those resources in each generation. Both Darwin and Wallace's understanding of this principle came from reading an essay by the economist Thomas Malthus who discussed this principle in relation to human populations. Third, offspring vary among each other in regard to their characteristics and those variations are inherited. Darwin and Wallace reasoned that offspring with inherited characteristics which allow them to best compete for limited resources will survive and have more offspring than those individuals with variations that are less able to compete. Because characteristics are inherited, these traits will be better represented in the next generation. This will lead to change in populations over generations in a process that Darwin called descent with modification. Ultimately, natural selection leads to greater adaptation of the population to its local environment; it is the only mechanism known for adaptive evolution.

Papers by Darwin and Wallace ([\[link\]](#)) presenting the idea of natural selection were read together in 1858 before the Linnean Society in London. The following year Darwin's book, *On the Origin of Species*, was published. His book outlined in considerable detail his arguments for evolution by natural selection.



(a)



(b)

Both (a) Charles Darwin and (b) Alfred Wallace wrote scientific papers on natural selection that were presented together before the Linnean Society in 1858.

Demonstrations of evolution by natural selection are time consuming and difficult to obtain. One of the best examples has been demonstrated in the very birds that helped to inspire Darwin's theory: the Galápagos finches. Peter and Rosemary Grant and their colleagues have studied Galápagos finch populations every year since 1976 and have provided important demonstrations of natural selection. The Grants found changes from one generation to the next in the distribution of beak shapes with the medium ground finch on the Galápagos island of Daphne Major. The birds have inherited variation in the bill shape with some birds having wide deep bills and others having thinner bills. During a period in which rainfall was higher than normal because of an El Niño, the large hard seeds that large-billed birds ate were reduced in number; however, there was an abundance of the small soft seeds which the small-billed birds ate. Therefore, survival and reproduction were much better in the following years for the small-billed birds. In the years following this El Niño, the Grants measured beak sizes in the population and found that the average bill size was smaller. Since bill size is an inherited trait, parents with smaller bills had more offspring and

the size of bills had evolved to be smaller. As conditions improved in 1987 and larger seeds became more available, the trend toward smaller average bill size ceased.

**Note:**

**Career Connection**

**Field Biologist**

Many people hike, explore caves, scuba dive, or climb mountains for recreation. People often participate in these activities hoping to see wildlife. Experiencing the outdoors can be incredibly enjoyable and invigorating. What if your job was to be outside in the wilderness? Field biologists by definition work outdoors in the “field.” The term field in this case refers to any location outdoors, even under water. A field biologist typically focuses research on a certain species, group of organisms, or a single habitat ([\[link\]](#)).



A field biologist tranquilizes a polar bear for study. (credit: Karen Rhode)

One objective of many field biologists includes discovering new species that have never been recorded. Not only do such findings expand our understanding of the natural world, but they also lead to important innovations in fields such as medicine and agriculture. Plant and microbial

species, in particular, can reveal new medicinal and nutritive knowledge. Other organisms can play key roles in ecosystems or be considered rare and in need of protection. When discovered, these important species can be used as evidence for environmental regulations and laws.

## Processes and Patterns of Evolution

Natural selection can only take place if there is **variation**, or differences, among individuals in a population. Importantly, these differences must have some genetic basis; otherwise, the selection will not lead to change in the next generation. This is critical because variation among individuals can be caused by non-genetic reasons such as an individual being taller because of better nutrition rather than different genes.

Genetic diversity in a population comes from two main mechanisms: mutation and sexual reproduction. Mutation, a change in DNA, is the ultimate source of new alleles, or new genetic variation in any population. The genetic changes caused by mutation can have one of three outcomes on the phenotype. A mutation affects the phenotype of the organism in a way that gives it reduced fitness—lower likelihood of survival or fewer offspring. A mutation may produce a phenotype with a beneficial effect on fitness. And, many mutations will also have no effect on the fitness of the phenotype; these are called neutral mutations. Mutations may also have a whole range of effect sizes on the fitness of the organism that expresses them in their phenotype, from a small effect to a great effect. Sexual reproduction also leads to genetic diversity: when two parents reproduce, unique combinations of alleles assemble to produce the unique genotypes and thus phenotypes in each of the offspring.

A heritable trait that helps the survival and reproduction of an organism in its present environment is called an **adaptation**. Scientists describe groups of organisms becoming adapted to their environment when a change in the range of genetic variation occurs over time that increases or maintains the “fit” of the population to its environment. The webbed feet of platypuses are an adaptation for swimming. The snow leopards’ thick fur is an



adaptation for living in the cold. The cheetahs' fast speed is an adaptation for catching prey.

Whether or not a trait is favorable depends on the environmental conditions at the time. The same traits are not always selected because environmental conditions can change. For example, consider a species of plant that grew in a moist climate and did not need to conserve water. Large leaves were selected because they allowed the plant to obtain more energy from the sun. Large leaves require more water to maintain than small leaves, and the moist environment provided favorable conditions to support large leaves. After thousands of years, the climate changed, and the area no longer had excess water. The direction of natural selection shifted so that plants with small leaves were selected because those populations were able to conserve water to survive the new environmental conditions.

The evolution of species has resulted in enormous variation in form and function. Sometimes, evolution gives rise to groups of organisms that become tremendously different from each other. When two species evolve in diverse directions from a common point, it is called **divergent evolution**. Such divergent evolution can be seen in the forms of the reproductive organs of flowering plants which share the same basic anatomies; however, they can look very different as a result of selection in different physical environments and adaptation to different kinds of pollinators ([\[link\]](#)).



Flowering plants evolved from a common ancestor. Notice that the (a) dense blazing

star (*Liatrus spicata*) and the (b) purple coneflower (*Echinacea purpurea*) vary in appearance, yet both share a similar basic morphology. (credit a: modification of work by Drew Avery; credit b: modification of work by Cory Zanker)

In other cases, similar phenotypes evolve independently in distantly related species. For example, flight has evolved in both bats and insects, and they both have structures we refer to as wings, which are adaptations to flight. However, the wings of bats and insects have evolved from very different original structures. This phenomenon is called **convergent evolution**, where similar traits evolve independently in species that do not share a common ancestry. The two species came to the same function, flying, but did so separately from each other.

These physical changes occur over enormous spans of time and help explain how evolution occurs. Natural selection acts on individual organisms, which in turn can shape an entire species. Although natural selection may work in a single generation on an individual, it can take thousands or even millions of years for the genotype of an entire species to evolve. It is over these large time spans that life on earth has changed and continues to change.

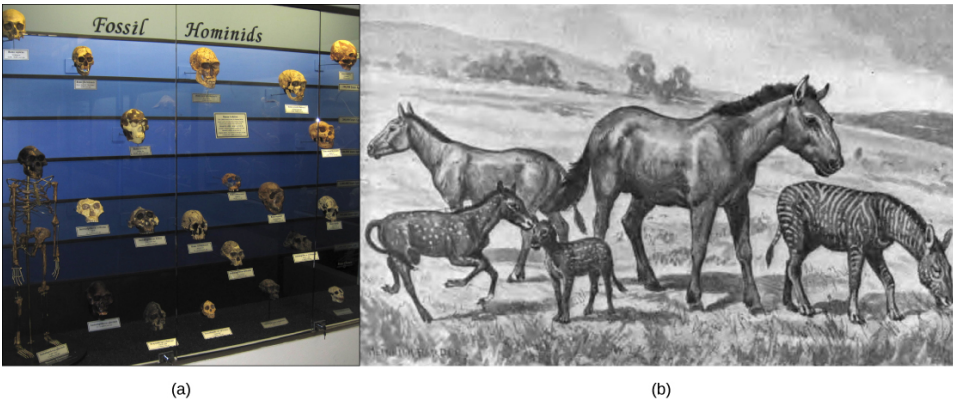
## Evidence of Evolution

The evidence for evolution is compelling and extensive. Looking at every level of organization in living systems, biologists see the signature of past and present evolution. Darwin dedicated a large portion of his book, *On the Origin of Species*, to identifying patterns in nature that were consistent with evolution, and since Darwin, our understanding has become clearer and broader.



## Fossils

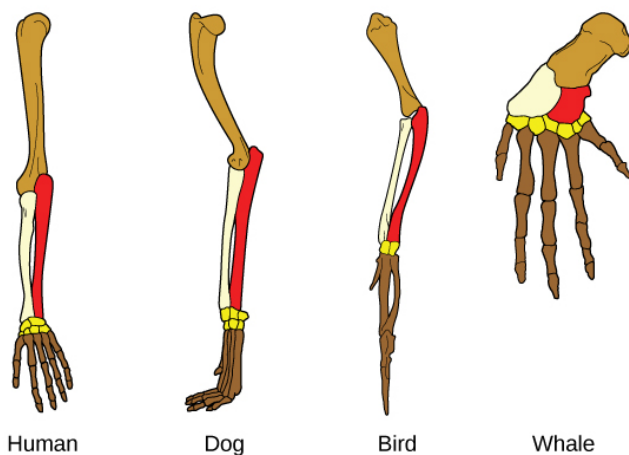
Fossils provide solid evidence that organisms from the past are not the same as those found today, and fossils show a progression of evolution. Scientists determine the age of fossils and categorize them from all over the world to determine when the organisms lived relative to each other. The resulting fossil record tells the story of the past and shows the evolution of form over millions of years ([\[link\]](#)). For example, scientists have recovered highly detailed records showing the evolution of humans and horses ([\[link\]](#)). The whale flipper shares a similar morphology to appendages of birds and mammals ([\[link\]](#)) indicating that these species share a common ancestor.



In this (a) display, fossil hominids are arranged from oldest (bottom) to newest (top). As hominids evolved, the shape of the skull changed. An artist's rendition of (b) extinct species of the genus *Equus* reveals that these ancient species resembled the modern horse (*Equus ferus*) but varied in size.

## Anatomy and Embryology

Another type of evidence for evolution is the presence of structures in organisms that share the same basic form. For example, the bones in the appendages of a human, dog, bird, and whale all share the same overall construction ([link](#)) resulting from their origin in the appendages of a common ancestor. Over time, evolution led to changes in the shapes and sizes of these bones in different species, but they have maintained the same overall layout. Scientists call these synonymous parts **homologous structures**.



The similar construction of these appendages indicates that these organisms share a common ancestor.

Some structures exist in organisms that have no apparent function at all, and appear to be residual parts from a past common ancestor. These unused structures without function are called **vestigial structures**. Other examples of vestigial structures are wings on flightless birds, leaves on some cacti, and hind leg bones in whales.

**Note:**

## Link to Learning



Visit this [interactive site](#) to guess which bones structures are homologous and which are analogous, and see examples of evolutionary adaptations to illustrate these concepts.

Another evidence of evolution is the convergence of form in organisms that share similar environments. For example, species of unrelated animals, such as the arctic fox and ptarmigan, living in the arctic region have been selected for seasonal white phenotypes during winter to blend with the snow and ice ([link](#)ab). These similarities occur not because of common ancestry, but because of similar selection pressures—the benefits of not being seen by predators.



(a)

(b)

The white winter coat of the (a) arctic fox

and the (b) ptarmigan's plumage are adaptations to their environments. (credit a: modification of work by Keith Morehouse)

Embryology, the study of the development of the anatomy of an organism to its adult form, also provides evidence of relatedness between now widely divergent groups of organisms. Mutational tweaking in the embryo can have such magnified consequences in the adult that embryo formation tends to be conserved. As a result, structures that are absent in some groups often appear in their embryonic forms and disappear by the time the adult or juvenile form is reached. For example, all vertebrate embryos, including humans, exhibit gill slits and tails at some point in their early development. These disappear in the adults of terrestrial groups but are maintained in adult forms of aquatic groups such as fish and some amphibians. Great ape embryos, including humans, have a tail structure during their development that is lost by the time of birth.

## **Biogeography**

The geographic distribution of organisms on the planet follows patterns that are best explained by evolution in conjunction with the movement of tectonic plates over geological time. Broad groups that evolved before the breakup of the supercontinent Pangaea (about 200 million years ago) are distributed worldwide. Groups that evolved since the breakup appear uniquely in regions of the planet, such as the unique flora and fauna of northern continents that formed from the supercontinent Laurasia and of the southern continents that formed from the supercontinent Gondwana. The presence of members of the plant family Proteaceae in Australia, southern Africa, and South America is best by their presence prior to the southern supercontinent Gondwana breaking up.

The great diversification of marsupials in Australia and the absence of other mammals reflect Australia's long isolation. Australia has an abundance of endemic species—species found nowhere else—which is typical of islands

whose isolation by expanses of water prevents species to migrate. Over time, these species diverge evolutionarily into new species that look very different from their ancestors that may exist on the mainland. The marsupials of Australia, the finches on the Galápagos, and many species on the Hawaiian Islands are all unique to their one point of origin, yet they display distant relationships to ancestral species on mainlands.

## **Molecular Biology**

Like anatomical structures, the structures of the molecules of life reflect descent with modification. Evidence of a common ancestor for all of life is reflected in the universality of DNA as the genetic material and in the near universality of the genetic code and the machinery of DNA replication and expression. Fundamental divisions in life between the three domains are reflected in major structural differences in otherwise conservative structures such as the components of ribosomes and the structures of membranes. In general, the relatedness of groups of organisms is reflected in the similarity of their DNA sequences—exactly the pattern that would be expected from descent and diversification from a common ancestor.

DNA sequences have also shed light on some of the mechanisms of evolution. For example, it is clear that the evolution of new functions for proteins commonly occurs after gene duplication events that allow the free modification of one copy by mutation, selection, or drift (changes in a population's gene pool resulting from chance), while the second copy continues to produce a functional protein.

## **Misconceptions of Evolution**

Although the theory of evolution generated some controversy when it was first proposed, it was almost universally accepted by biologists, particularly younger biologists, within 20 years after publication of *On the Origin of Species*. Nevertheless, the theory of evolution is a difficult concept and misconceptions about how it works abound.

**Note:****Link to Learning**

This [site](#) addresses some of the main misconceptions associated with the theory of evolution.

## **Evolution Is Just a Theory**

Critics of the theory of evolution dismiss its importance by purposefully confounding the everyday usage of the word “theory” with the way scientists use the word. In science, a “theory” is understood to be a body of thoroughly tested and verified explanations for a set of observations of the natural world. Scientists have a theory of the atom, a theory of gravity, and the theory of relativity, each of which describes understood facts about the world. In the same way, the theory of evolution describes facts about the living world. As such, a theory in science has survived significant efforts to discredit it by scientists. In contrast, a “theory” in common vernacular is a word meaning a guess or suggested explanation; this meaning is more akin to the scientific concept of “hypothesis.” When critics of evolution say evolution is “just a theory,” they are implying that there is little evidence supporting it and that it is still in the process of being rigorously tested. This is a mischaracterization.

## **Individuals Evolve**

Evolution is the change in genetic composition of a population over time, specifically over generations, resulting from differential reproduction of individuals with certain alleles. Individuals do change over their lifetime,

obviously, but this is called development and involves changes programmed by the set of genes the individual acquired at birth in coordination with the individual's environment. When thinking about the evolution of a characteristic, it is probably best to think about the change of the average value of the characteristic in the population over time. For example, when natural selection leads to bill-size change in medium-ground finches in the Galápagos, this does not mean that individual bills on the finches are changing. If one measures the average bill size among all individuals in the population at one time and then measures the average bill size in the population several years later, this average value will be different as a result of evolution. Although some individuals may survive from the first time to the second, they will still have the same bill size; however, there will be many new individuals that contribute to the shift in average bill size.

## **Evolution Explains the Origin of Life**

It is a common misunderstanding that evolution includes an explanation of life's origins. Conversely, some of the theory's critics believe that it cannot explain the origin of life. The theory does not try to explain the origin of life. The theory of evolution explains how populations change over time and how life diversifies the origin of species. It does not shed light on the beginnings of life including the origins of the first cells, which is how life is defined. The mechanisms of the origin of life on Earth are a particularly difficult problem because it occurred a very long time ago, and presumably it just occurred once. Importantly, biologists believe that the presence of life on Earth precludes the possibility that the events that led to life on Earth can be repeated because the intermediate stages would immediately become food for existing living things.

However, once a mechanism of inheritance was in place in the form of a molecule like DNA either within a cell or pre-cell, these entities would be subject to the principle of natural selection. More effective reproducers would increase in frequency at the expense of inefficient reproducers. So while evolution does not explain the origin of life, it may have something to say about some of the processes operating once pre-living entities acquired certain properties.

## Organisms Evolve on Purpose

Statements such as “organisms evolve in response to a change in an environment” are quite common, but such statements can lead to two types of misunderstandings. First, the statement must not be understood to mean that individual organisms evolve. The statement is shorthand for “a population evolves in response to a changing environment.” However, a second misunderstanding may arise by interpreting the statement to mean that the evolution is somehow intentional. A changed environment results in some individuals in the population, those with particular phenotypes, benefiting and therefore producing proportionately more offspring than other phenotypes. This results in change in the population if the characteristics are genetically determined.

It is also important to understand that the variation that natural selection works on is already in a population and does not arise in response to an environmental change. For example, applying antibiotics to a population of bacteria will, over time, select a population of bacteria that are resistant to antibiotics. The resistance, which is caused by a gene, did not arise by mutation because of the application of the antibiotic. The gene for resistance was already present in the gene pool of the bacteria, likely at a low frequency. The antibiotic, which kills the bacterial cells without the resistance gene, strongly selects individuals that are resistant, since these would be the only ones that survived and divided. Experiments have demonstrated that mutations for antibiotic resistance do not arise as a result of antibiotic.

In a larger sense, evolution is not goal directed. Species do not become “better” over time; they simply track their changing environment with adaptations that maximize their reproduction in a particular environment at a particular time. Evolution has no goal of making faster, bigger, more complex, or even smarter species, despite the commonness of this kind of language in popular discourse. What characteristics evolve in a species are a function of the variation present and the environment, both of which are constantly changing in a non-directional way. What trait is fit in one environment at one time may well be fatal at some point in the future. This holds equally well for a species of insect as it does the human species.



## Section Summary

Evolution is the process of adaptation through mutation which allows more desirable characteristics to be passed to the next generation. Over time, organisms evolve more characteristics that are beneficial to their survival. For living organisms to adapt and change to environmental pressures, genetic variation must be present. With genetic variation, individuals have differences in form and function that allow some to survive certain conditions better than others. These organisms pass their favorable traits to their offspring. Eventually, environments change, and what was once a desirable, advantageous trait may become an undesirable trait and organisms may further evolve. Evolution may be convergent with similar traits evolving in multiple species or divergent with diverse traits evolving in multiple species that came from a common ancestor. Evidence of evolution can be observed by means of DNA code and the fossil record, and also by the existence of homologous and vestigial structures.

## Review Questions

### Exercise:

#### Problem:

Which scientific concept did Charles Darwin and Alfred Wallace independently discover?

- a. mutation
- b. natural selection
- c. overbreeding
- d. sexual reproduction

---

#### Solution:

B

### Exercise:

**Problem:**

Which of the following situations will lead to natural selection?

- a. The seeds of two plants land near each other and one grows larger than the other.
- b. Two types of fish eat the same kind of food, and one is better able to gather food than the other.
- c. Male lions compete for the right to mate with females, with only one possible winner.
- d. all of the above

---

**Solution:**

D

**Exercise:**

**Problem:** Which description is an example of a phenotype?

- a. A certain duck has a blue beak.
- b. A mutation occurred to a flower.
- c. Most cheetahs live solitary lives.
- d. both a and c

---

**Solution:**

D

**Exercise:****Problem:**

Which situation is most likely an example of convergent evolution?

- a. Squid and humans have eyes similar in structure.
- b. Worms and snakes both move without legs.

- c. Some bats and birds have wings that allow them to fly
- d. all of the above

---

**Solution:**

D

## **Free Response**

### **Exercise:**

**Problem:**

If a person scatters a handful of garden pea plant seeds in one area, how would natural selection work in this situation?

---

**Solution:**

The plants that can best use the resources of the area, including competing with other individuals for those resources will produce more seeds themselves and those traits that allowed them to better use the resources will increase in the population of the next generation.

### **Exercise:**

**Problem:**

Why do scientists consider vestigial structures evidence for evolution?

---

**Solution:**

Vestigial structures are considered evidence for evolution because most structures do not exist in an organism without serving some function either presently or in the past. A vestigial structure indicates a past form or function that has since changed, but the structure remains present because it had a function in the ancestor.

### **Exercise:**

**Problem:**

How does the scientific meaning of “theory” differ from the common vernacular meaning?

---

**Solution:**

In science, a theory is a thoroughly tested and verified set of explanations for a body of observations of nature. It is the strongest form of knowledge in science. In contrast, a theory in common vernacular can mean a guess or speculation about something, meaning that the knowledge implied by the theory is very weak.

**Exercise:****Problem:**

Explain why the statement that a monkey is more evolved than a mouse is incorrect.

---

**Solution:**

The statement implies that there is a goal to evolution and that the monkey represents greater progress to that goal than the mouse. Both species are likely to be well adapted to their particular environments, which is the outcome of natural selection.

**Glossary****adaptation**

heritable trait or behavior in an organism that aids in its survival and reproduction in its present environment

**convergent evolution**

process by which groups of organisms independently evolve to similar forms

**divergent evolution**

process by which groups of organisms evolve in diverse directions from a common point

homologous structures

parallel structures in diverse organisms that have a common ancestor

natural selection

reproduction of individuals with favorable genetic traits that survive environmental change because of those traits, leading to evolutionary change

variation

genetic differences among individuals in a population

vestigial structure

physical structure present in an organism but that has no apparent function and appears to be from a functional structure in a distant ancestor

## Formation of New Species

By the end of this section, you will be able to:

- Define species and describe how species are identified as different
- Describe genetic variables that lead to speciation
- Identify prezygotic and postzygotic reproductive barriers
- Explain allopatric and sympatric speciation
- Describe adaptive radiation

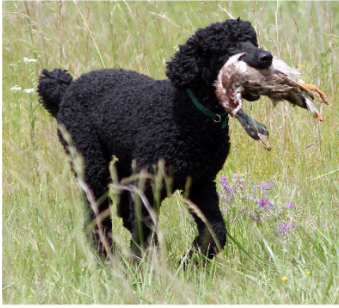
Although all life on earth shares various genetic similarities, only certain organisms combine genetic information by sexual reproduction and have offspring that can then successfully reproduce. Scientists call such organisms members of the same biological species.

## Species and the Ability to Reproduce

A **species** is a group of individual organisms that interbreed and produce fertile, viable offspring. According to this definition, one species is distinguished from another when, in nature, it is not possible for matings between individuals from each species to produce fertile offspring.

Members of the same species share both external and internal characteristics, which develop from their DNA. The closer relationship two organisms share, the more DNA they have in common, just like people and their families. People's DNA is likely to be more like their father or mother's DNA than their cousin or grandparent's DNA. Organisms of the same species have the highest level of DNA alignment and therefore share characteristics and behaviors that lead to successful reproduction.

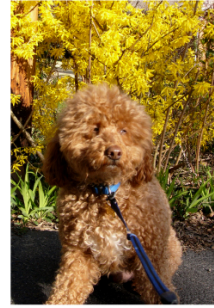
Species' appearance can be misleading in suggesting an ability or inability to mate. For example, even though domestic dogs (*Canis lupus familiaris*) display phenotypic differences, such as size, build, and coat, most dogs can interbreed and produce viable puppies that can mature and sexually reproduce ([link](#)).



(a)



(b)



(c)

The (a) poodle and (b) cocker spaniel can reproduce to produce a breed known as (c) the cockapoo. (credit a: modification of work by Sally Eller, Tom Reese; credit b: modification of work by Jeremy McWilliams; credit c: modification of work by Kathleen Conklin)

In other cases, individuals may appear similar although they are not members of the same species. For example, even though bald eagles (*Haliaeetus leucocephalus*) and African fish eagles (*Haliaeetus vocifer*) are both birds and eagles, each belongs to a separate species group ([\[link\]](#)). If humans were to artificially intervene and fertilize the egg of a bald eagle with the sperm of an African fish eagle and a chick did hatch, that offspring, called a **hybrid** (a cross between two species), would probably be infertile—unable to successfully reproduce after it reached maturity. Different species may have different genes that are active in development; therefore, it may not be possible to develop a viable offspring with two different sets of directions. Thus, even though hybridization may take place, the two species still remain separate.



(a)



(b)

The (a) African fish eagle is similar in appearance to the (b) bald eagle, but the two birds are members of different species.  
(credit a: modification of work by Nigel Wedge; credit b: modification of work by U.S. Fish and Wildlife Service)

Populations of species share a gene pool: a collection of all the variants of genes in the species. Again, the basis to any changes in a group or population of organisms must be genetic for this is the only way to share and pass on traits. When variations occur within a species, they can only be passed to the next generation along two main pathways: asexual reproduction or sexual reproduction. The change will be passed on asexually simply if the reproducing cell possesses the changed trait. For the changed trait to be passed on by sexual reproduction, a gamete, such as a sperm or egg cell, must possess the changed trait. In other words, sexually-reproducing organisms can experience several genetic changes in their body cells, but if these changes do not occur in a sperm or egg cell, the changed trait will never reach the next generation. Only heritable traits can evolve. Therefore, reproduction plays a paramount role for genetic change to take root in a population or species. In short, organisms must be able to reproduce with each other to pass new traits to offspring.

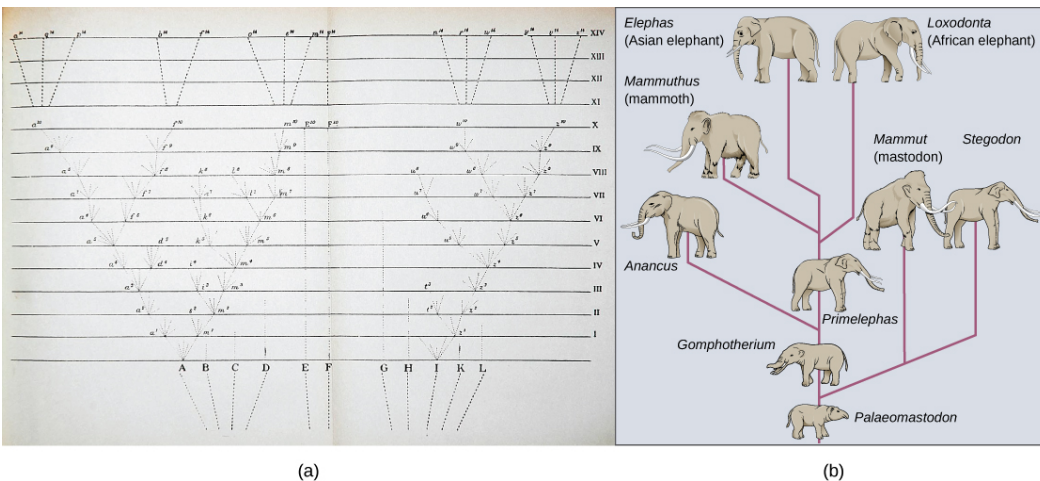
## Speciation

The biological definition of species, which works for sexually reproducing organisms, is a group of actually or potentially interbreeding individuals.



There are exceptions to this rule. Many species are similar enough that hybrid offspring are possible and may often occur in nature, but for the majority of species this rule generally holds. In fact, the presence in nature of hybrids between similar species suggests that they may have descended from a single interbreeding species, and the speciation process may not yet be completed.

Given the extraordinary diversity of life on the planet there must be mechanisms for **speciation**: the formation of two species from one original species. Darwin envisioned this process as a branching event and diagrammed the process in the only illustration found in *On the Origin of Species* ([link](#)[a](#)). Compare this illustration to the diagram of elephant evolution ([link](#)[b](#)), which shows that as one species changes over time, it branches to form more than one new species, repeatedly, as long as the population survives or until the organism becomes extinct.



The only illustration in Darwin's *On the Origin of Species* is (a) a diagram showing speciation events leading to biological diversity. The diagram shows similarities to phylogenetic charts that are drawn today to illustrate the relationships of species. (b) Modern elephants evolved from the *Palaeomastodon*, a species that lived in Egypt 35–50 million years ago.

For speciation to occur, two new populations must be formed from one original population and they must evolve in such a way that it becomes impossible for individuals from the two new populations to interbreed. Biologists have proposed mechanisms by which this could occur that fall into two broad categories. **Allopatric speciation** (allo- = "other"; -patric = "homeland") involves geographic separation of populations from a parent species and subsequent evolution. **Sympatric speciation** (sym- = "same"; -patric = "homeland") involves speciation occurring within a parent species remaining in one location.

Biologists think of speciation events as the splitting of one ancestral species into two descendant species. There is no reason why there might not be more than two species formed at one time except that it is less likely and multiple events can be conceptualized as single splits occurring close in time.

## **Allopatric Speciation**

A geographically continuous population has a gene pool that is relatively homogeneous. Gene flow, the movement of alleles across the range of the species, is relatively free because individuals can move and then mate with individuals in their new location. Thus, the frequency of an allele at one end of a distribution will be similar to the frequency of the allele at the other end. When populations become geographically discontinuous, that free-flow of alleles is prevented. When that separation lasts for a period of time, the two populations are able to evolve along different trajectories. Thus, their allele frequencies at numerous genetic loci gradually become more and more different as new alleles independently arise by mutation in each population. Typically, environmental conditions, such as climate, resources, predators, and competitors for the two populations will differ causing natural selection to favor divergent adaptations in each group.

Isolation of populations leading to allopatric speciation can occur in a variety of ways: a river forming a new branch, erosion forming a new valley, a group of organisms traveling to a new location without the ability to return, or seeds floating over the ocean to an island. The nature of the

geographic separation necessary to isolate populations depends entirely on the biology of the organism and its potential for dispersal. If two flying insect populations took up residence in separate nearby valleys, chances are, individuals from each population would fly back and forth continuing gene flow. However, if two rodent populations became divided by the formation of a new lake, continued gene flow would be unlikely; therefore, speciation would be more likely.

Biologists group allopatric processes into two categories: dispersal and vicariance. **Dispersal** is when a few members of a species move to a new geographical area, and **vicariance** is when a natural situation arises to physically divide organisms.

Scientists have documented numerous cases of allopatric speciation taking place. For example, along the west coast of the United States, two separate sub-species of spotted owls exist. The northern spotted owl has genetic and phenotypic differences from its close relative: the Mexican spotted owl, which lives in the south ([link](#)).



The northern spotted owl and

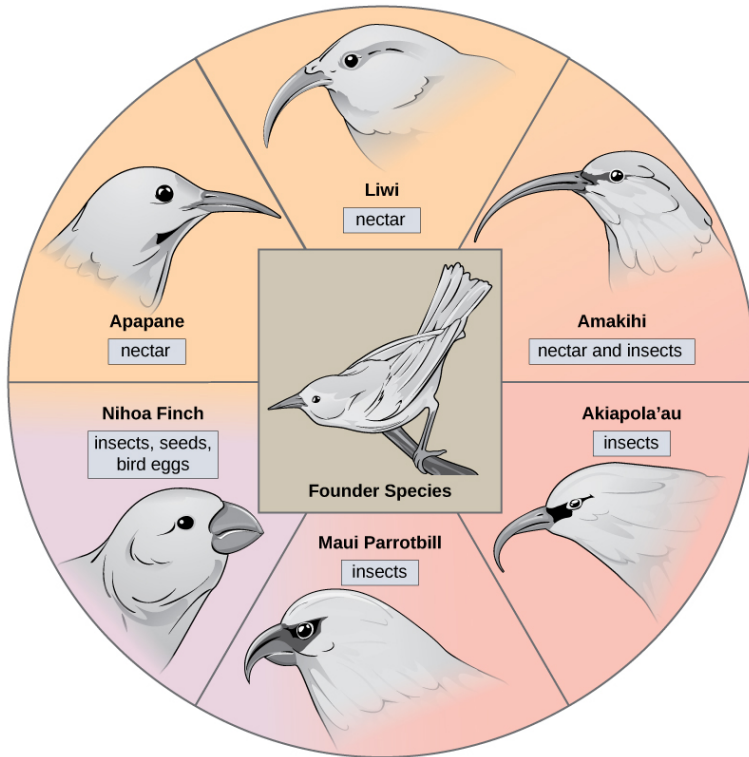
the Mexican spotted owl inhabit geographically separate locations with different climates and ecosystems. The owl is an example of allopatric speciation. (credit "northern spotted owl": modification of work by John and Karen Hollingsworth; credit "Mexican spotted owl": modification of work by Bill Radke)

Additionally, scientists have found that the further the distance between two groups that once were the same species, the more likely it is that speciation will occur. This seems logical because as the distance increases, the various environmental factors would likely have less in common than locations in close proximity. Consider the two owls: in the north, the climate is cooler than in the south; the types of organisms in each ecosystem differ, as do their behaviors and habits; also, the hunting habits and prey choices of the southern owls vary from the northern owls. These variances can lead to evolved differences in the owls, and speciation likely will occur.

## **Adaptive Radiation**

In some cases, a population of one species disperses throughout an area, and each finds a distinct niche or isolated habitat. Over time, the varied demands of their new lifestyles lead to multiple speciation events originating from a single species. This is called **adaptive radiation** because many adaptations evolve from a single point of origin; thus, causing the species to radiate into several new ones. Island archipelagos like the Hawaiian Islands provide an ideal context for adaptive radiation events because water surrounds each island which leads to geographical isolation

for many organisms. The Hawaiian honeycreeper illustrates one example of adaptive radiation. From a single species, called the founder species, numerous species have evolved, including the six shown in [\[link\]](#).



The honeycreeper birds illustrate adaptive radiation. From one original species of bird, multiple others evolved, each with its own distinctive characteristics.

Notice the differences in the species' beaks in [\[link\]](#). Evolution in response to natural selection based on specific food sources in each new habitat led to evolution of a different beak suited to the specific food source. The seed-eating bird has a thicker, stronger beak which is suited to break hard nuts. The nectar-eating birds have long beaks to dip into flowers to reach the nectar. The insect-eating birds have beaks like swords, appropriate for

stabbing and impaling insects. Darwin's finches are another example of adaptive radiation in an archipelago.

**Note:**

Link to Learning



Click through this [interactive site](#) to see how island birds evolved in evolutionary increments from 5 million years ago to today.

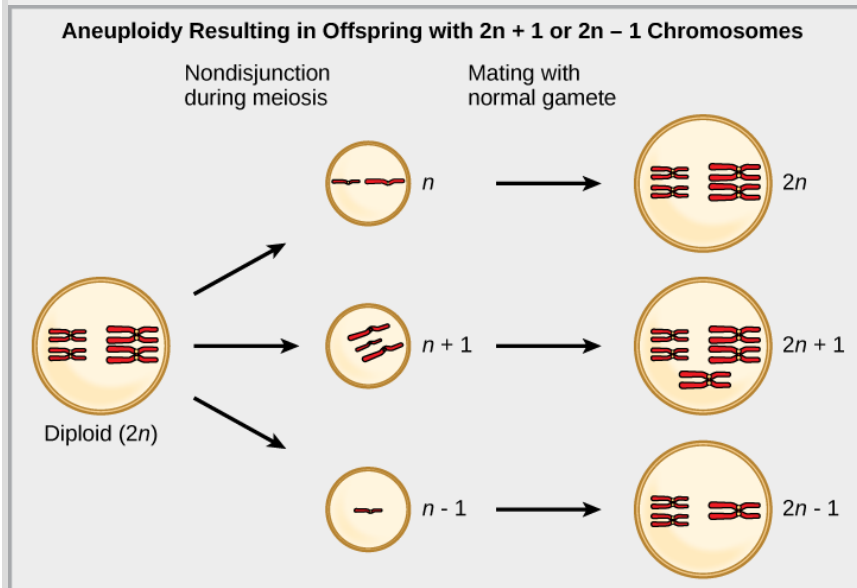
## Sympatric Speciation

Can divergence occur if no physical barriers are in place to separate individuals who continue to live and reproduce in the same habitat? The answer is yes. The process of speciation within the same space is called sympatric speciation; the prefix “sym” means same, so “sympatric” means “same homeland” in contrast to “allopatric” meaning “other homeland.” A number of mechanisms for sympatric speciation have been proposed and studied.

One form of sympatric speciation can begin with a serious chromosomal error during cell division. In a normal cell division event chromosomes replicate, pair up, and then separate so that each new cell has the same number of chromosomes. However, sometimes the pairs separate and the end cell product has too many or too few individual chromosomes in a condition called **aneuploidy** ([link](#)).

## Note:

### Art Connection

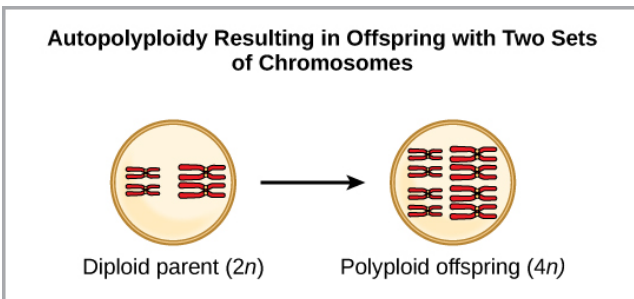


Aneuploidy results when the gametes have too many or too few chromosomes due to nondisjunction during meiosis. In the example shown here, the resulting offspring will have  $2n+1$  or  $2n-1$  chromosomes

Which is most likely to survive, offspring with  $2n+1$  chromosomes or offspring with  $2n-1$  chromosomes?

Polyploidy is a condition in which a cell or organism has an extra set, or sets, of chromosomes. Scientists have identified two main types of polyploidy that can lead to reproductive isolation of an individual in the polyploidy state. Reproductive isolation is the inability to interbreed. In some cases, a polyploid individual will have two or more complete sets of chromosomes from its own species in a condition called **autopolyploidy** ([link](#)). The prefix “auto-” means “self,” so the term means multiple chromosomes from one’s own species. Polyploidy results from an error in

meiosis in which all of the chromosomes move into one cell instead of separating.

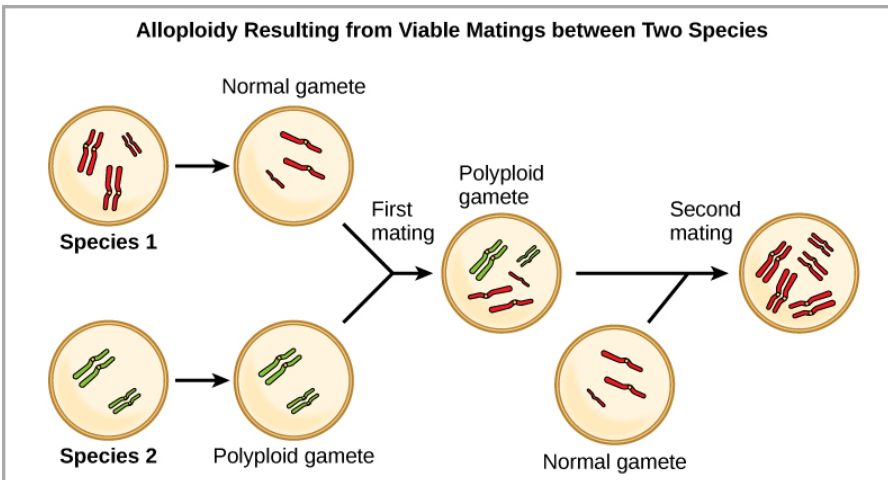


Autopolyploidy results when  
mitosis is not followed by  
cytokinesis.

For example, if a plant species with  $2n = 6$  produces autopolyploid gametes that are also diploid ( $2n = 6$ , when they should be  $n = 3$ ), the gametes now have twice as many chromosomes as they should have. These new gametes will be incompatible with the normal gametes produced by this plant species. However, they could either self-pollinate or reproduce with other autopolyploid plants with gametes having the same diploid number. In this way, sympatric speciation can occur quickly by forming offspring with  $4n$  called a tetraploid. These individuals would immediately be able to reproduce only with those of this new kind and not those of the ancestral species.

The other form of polyploidy occurs when individuals of two different species reproduce to form a viable offspring called an **allopolyploid**. The prefix “allo-” means “other” (recall from allopatric): therefore, an allopolyploid occurs when gametes from two different species combine. [\[link\]](#) illustrates one possible way an allopolyploid can form. Notice how it takes two generations, or two reproductive acts, before the viable fertile hybrid results.





Allopolyploidy results when two species mate to produce viable offspring. In the example shown, a normal gamete from one species fuses with a polyploid gamete from another. Two matings are necessary to produce viable offspring.

The cultivated forms of wheat, cotton, and tobacco plants are all allopolyploids. Although polyploidy occurs occasionally in animals, it takes place most commonly in plants. (Animals with any of the types of chromosomal aberrations described here are unlikely to survive and produce normal offspring.) Scientists have discovered more than half of all plant species studied relate back to a species evolved through polyploidy. With such a high rate of polyploidy in plants, some scientists hypothesize that this mechanism takes place more as an adaptation than as an error.

## Reproductive Isolation

Given enough time, the genetic and phenotypic divergence between populations will affect characters that influence reproduction: if individuals of the two populations were to be brought together, mating would be less likely, but if mating occurred, offspring would be non-viable or infertile. Many types of diverging characters may affect the **reproductive isolation**, the ability to interbreed, of the two populations.

Reproductive isolation can take place in a variety of ways. Scientists organize them into two groups: prezygotic barriers and postzygotic barriers. Recall that a zygote is a fertilized egg; the first cell of the development of an organism that reproduces sexually. Therefore, a **prezygotic barrier** is a mechanism that blocks reproduction from taking place; this includes barriers that prevent fertilization when organisms attempt reproduction. A **postzygotic barrier** occurs after zygote formation; this includes organisms that don't survive the embryonic stage and those that are born sterile.

Some types of prezygotic barriers prevent reproduction entirely. Many organisms only reproduce at certain times of the year, often just annually. Differences in breeding schedules, called **temporal isolation**, can act as a form of reproductive isolation. For example, two species of frogs inhabit the same area, but one reproduces from January to March, whereas the other reproduces from March to May ([link](#)).



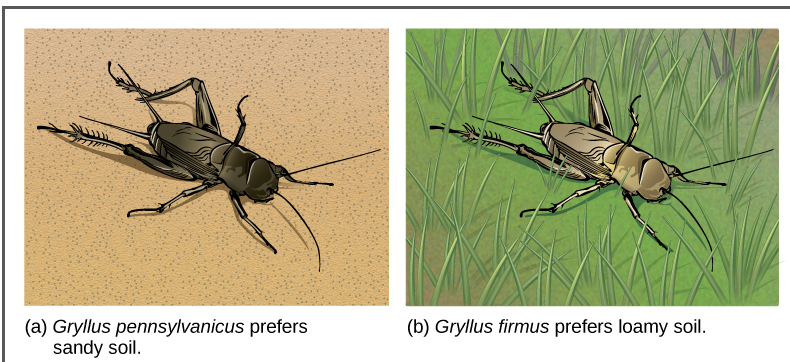
(a)

(b)

These two related frog species exhibit temporal reproductive isolation. (a) *Rana aurora* breeds earlier in the year than (b) *Rana boylei*. (credit a: modification of work by Mark R. Jennings, USFWS; credit b: modification of work by Alessandro Catenazzi)

In some cases, populations of a species move or are moved to a new habitat and take up residence in a place that no longer overlaps with the other

populations of the same species. This situation is called **habitat isolation**. Reproduction with the parent species ceases, and a new group exists that is now reproductively and genetically independent. For example, a cricket population that was divided after a flood could no longer interact with each other. Over time, the forces of natural selection, mutation, and genetic drift will likely result in the divergence of the two groups ([link](#)).



Speciation can occur when two populations occupy different habitats. The habitats need not be far apart. The cricket (a) *Gryllus pennsylvanicus* prefers sandy soil, and the cricket (b) *Gryllus firmus* prefers loamy soil. The two species can live in close proximity, but because of their different soil preferences, they became genetically isolated.

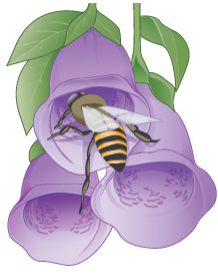
**Behavioral isolation** occurs when the presence or absence of a specific behavior prevents reproduction from taking place. For example, male fireflies use specific light patterns to attract females. Various species of fireflies display their lights differently. If a male of one species tried to attract the female of another, she would not recognize the light pattern and would not mate with the male.

Other prezygotic barriers work when differences in their gamete cells (eggs and sperm) prevent fertilization from taking place; this is called a **gametic barrier**. Similarly, in some cases closely related organisms try to mate, but their reproductive structures simply do not fit together. For example, damselfly males of different species have differently shaped reproductive organs. If one species tries to mate with the female of another, their body parts simply do not fit together. ([link](#)).



The shape of the male reproductive organ varies among male damselfly species, and is only compatible with the female of that species. Reproductive organ incompatibility keeps the species reproductively isolated.

In plants, certain structures aimed to attract one type of pollinator simultaneously prevent a different pollinator from accessing the pollen. The tunnel through which an animal must access nectar can vary widely in length and diameter, which prevents the plant from being cross-pollinated with a different species ([link](#)).



(a) Honeybee drinking nectar from a foxglove flower



(b) Ruby-throated hummingbird drinking nectar from a trumpet creeper flower

Some flowers have evolved to attract certain pollinators. The (a) wide foxglove flower is adapted for pollination by bees, while the (b) long, tube-shaped trumpet creeper flower is adapted for pollination by humming birds.

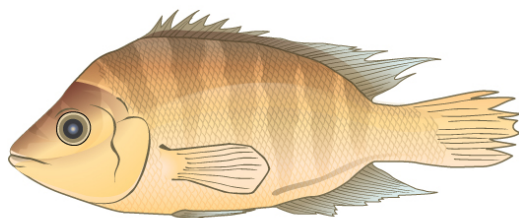
When fertilization takes place and a zygote forms, postzygotic barriers can prevent reproduction. Hybrid individuals in many cases cannot form normally in the womb and simply do not survive past the embryonic stages. This is called **hybrid inviability** because the hybrid organisms simply are not viable. In another postzygotic situation, reproduction leads to the birth and growth of a hybrid that is sterile and unable to reproduce offspring of their own; this is called hybrid sterility.

## Habitat Influence on Speciation

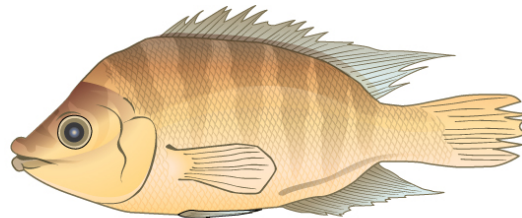
Sympatric speciation may also take place in ways other than polyploidy. For example, consider a species of fish that lives in a lake. As the population grows, competition for food also grows. Under pressure to find food, suppose that a group of these fish had the genetic flexibility to discover and feed off another resource that was unused by the other fish. What if this new food source was found at a different depth of the lake? Over time, those feeding on the second food source would interact more with each other than the other fish; therefore, they would breed together as well. Offspring of these fish would likely behave as their parents: feeding

and living in the same area and keeping separate from the original population. If this group of fish continued to remain separate from the first population, eventually sympatric speciation might occur as more genetic differences accumulated between them.

This scenario does play out in nature, as do others that lead to reproductive isolation. One such place is Lake Victoria in Africa, famous for its sympatric speciation of cichlid fish. Researchers have found hundreds of sympatric speciation events in these fish, which have not only happened in great number, but also over a short period of time. [\[link\]](#) shows this type of speciation among a cichlid fish population in Nicaragua. In this locale, two types of cichlids live in the same geographic location but have come to have different morphologies that allow them to eat various food sources.



Thin-lipped cichlid



Thick-lipped cichlid

Cichlid fish from Lake Apoyeque, Nicaragua, show evidence of sympatric speciation. Lake Apoyeque, a crater lake, is 1800 years old, but genetic evidence indicates that the lake was populated only 100 years ago by a single population of cichlid fish. Nevertheless, two populations with distinct morphologies and diets now exist in the lake, and scientists believe these populations may be in an early stage of speciation.

## Section Summary

Speciation occurs along two main pathways: geographic separation (allopatric speciation) and through mechanisms that occur within a shared habitat (sympatric speciation). Both pathways isolate a population

reproductively in some form. Mechanisms of reproductive isolation act as barriers between closely related species, enabling them to diverge and exist as genetically independent species. Prezygotic barriers block reproduction prior to formation of a zygote, whereas postzygotic barriers block reproduction after fertilization occurs. For a new species to develop, something must cause a breach in the reproductive barriers. Sympatric speciation can occur through errors in meiosis that form gametes with extra chromosomes (polyploidy). Autopolyploidy occurs within a single species, whereas allopolyploidy occurs between closely related species.

## Art Connections

### Exercise:

#### Problem:

[\[link\]](#) Which is most likely to survive, offspring with  $2n+1$  chromosomes or offspring with  $2n-1$  chromosomes?

---

#### Solution:

[\[link\]](#) Loss of genetic material is almost always lethal, so offspring with  $2n+1$  chromosomes are more likely to survive.

## Review Questions

### Exercise:

#### Problem:

Which situation would most likely lead to allopatric speciation?

- a. flood causes the formation of a new lake.
- b. A storm causes several large trees to fall down.
- c. A mutation causes a new trait to develop.
- d. An injury causes an organism to seek out a new food source.



---

**Solution:**

A

**Exercise:**

**Problem:**

What is the main difference between dispersal and vicariance?

- a. One leads to allopatric speciation, whereas the other leads to sympatric speciation.
- b. One involves the movement of the organism, and the other involves a change in the environment.
- c. One depends on a genetic mutation occurring, and the other does not.
- d. One involves closely related organisms, and the other involves only individuals of the same species.

---

**Solution:**

B

**Exercise:**

**Problem:**

Which variable increases the likelihood of allopatric speciation taking place more quickly?

- a. lower rate of mutation
- b. longer distance between divided groups
- c. increased instances of hybrid formation
- d. equivalent numbers of individuals in each population

---

**Solution:**

B



**Exercise:**

**Problem:**

What is the main difference between autopolyploid and allopolyploid?

- a. the number of chromosomes
- b. the functionality of the chromosomes
- c. the source of the extra chromosomes
- d. the number of mutations in the extra chromosomes

---

**Solution:**

C

**Exercise:**

**Problem:** Which reproductive combination produces hybrids?

- a. when individuals of the same species in different geographical areas reproduce
- b. when any two individuals sharing the same habitat reproduce
- c. when members of closely related species reproduce
- d. when offspring of the same parents reproduce

---

**Solution:**

C

**Exercise:**

**Problem:**

Which condition is the basis for a species to be reproductively isolated from other members?

- a. It does not share its habitat with related species.
- b. It does not exist out of a single habitat.

- c. It does not exchange genetic information with other species.
- d. It does not undergo evolutionary changes for a significant period of time.

---

**Solution:**

C

**Exercise:**

**Problem:** Which situation is *not* an example of a prezygotic barrier?

- a. Two species of turtles breed at different times of the year.
- b. Two species of flowers attract different pollinators.
- c. Two species of birds display different mating dances.
- d. Two species of insects produce infertile offspring.

---

**Solution:**

D

## Free Response

**Exercise:**

**Problem:**

Why do island chains provide ideal conditions for adaptive radiation to occur?

---

**Solution:**

Organisms of one species can arrive to an island together and then disperse throughout the chain, each settling into different niches and exploiting different food resources to reduce competition.

**Exercise:**

**Problem:**

Two species of fish had recently undergone sympatric speciation. The males of each species had a different coloring through which the females could identify and choose a partner from her own species. After some time, pollution made the lake so cloudy that it was hard for females to distinguish colors. What might take place in this situation?

---

**Solution:**

It is likely the two species would start to reproduce with each other. Depending on the viability of their offspring, they may fuse back into one species.

**Exercise:****Problem:**

Why can polyploidy individuals lead to speciation fairly quickly?

---

**Solution:**

The formation of gametes with new  $n$  numbers can occur in one generation. After a couple of generations, enough of these new hybrids can form to reproduce together as a new species.

**Glossary**

adaptive radiation

speciation when one species radiates out to form several other species

allopatric speciation

speciation that occurs via geographic separation

allopolyploid

polyploidy formed between two related, but separate species

aneuploidy

condition of a cell having an extra chromosome or missing a chromosome for its species

autopolyploid

polyploidy formed within a single species

behavioral isolation

type of reproductive isolation that occurs when a specific behavior or lack of one prevents reproduction from taking place

dispersal

allopatric speciation that occurs when a few members of a species move to a new geographical area

gametic barrier

prezygotic barrier occurring when closely related individuals of different species mate, but differences in their gamete cells (eggs and sperm) prevent fertilization from taking place

habitat isolation

reproductive isolation resulting when populations of a species move or are moved to a new habitat, taking up residence in a place that no longer overlaps with the other populations of the same species

hybrid

offspring of two closely related individuals, not of the same species

postzygotic barrier

reproductive isolation mechanism that occurs after zygote formation

prezygotic barrier

reproductive isolation mechanism that occurs before zygote formation

reproductive isolation

situation that occurs when a species is reproductively independent from other species; this may be brought about by behavior, location, or reproductive barriers

speciation

formation of a new species

species

group of populations that interbreed and produce fertile offspring

sympatric speciation

speciation that occurs in the same geographic space

temporal isolation

differences in breeding schedules that can act as a form of prezygotic barrier leading to reproductive isolation

vicariance

allopatric speciation that occurs when something in the environment separates organisms of the same species into separate groups

## Reconnection and Rates of Speciation

By the end of this section, you will be able to:

- Describe pathways of species evolution in hybrid zones
- Explain the two major theories on rates of speciation

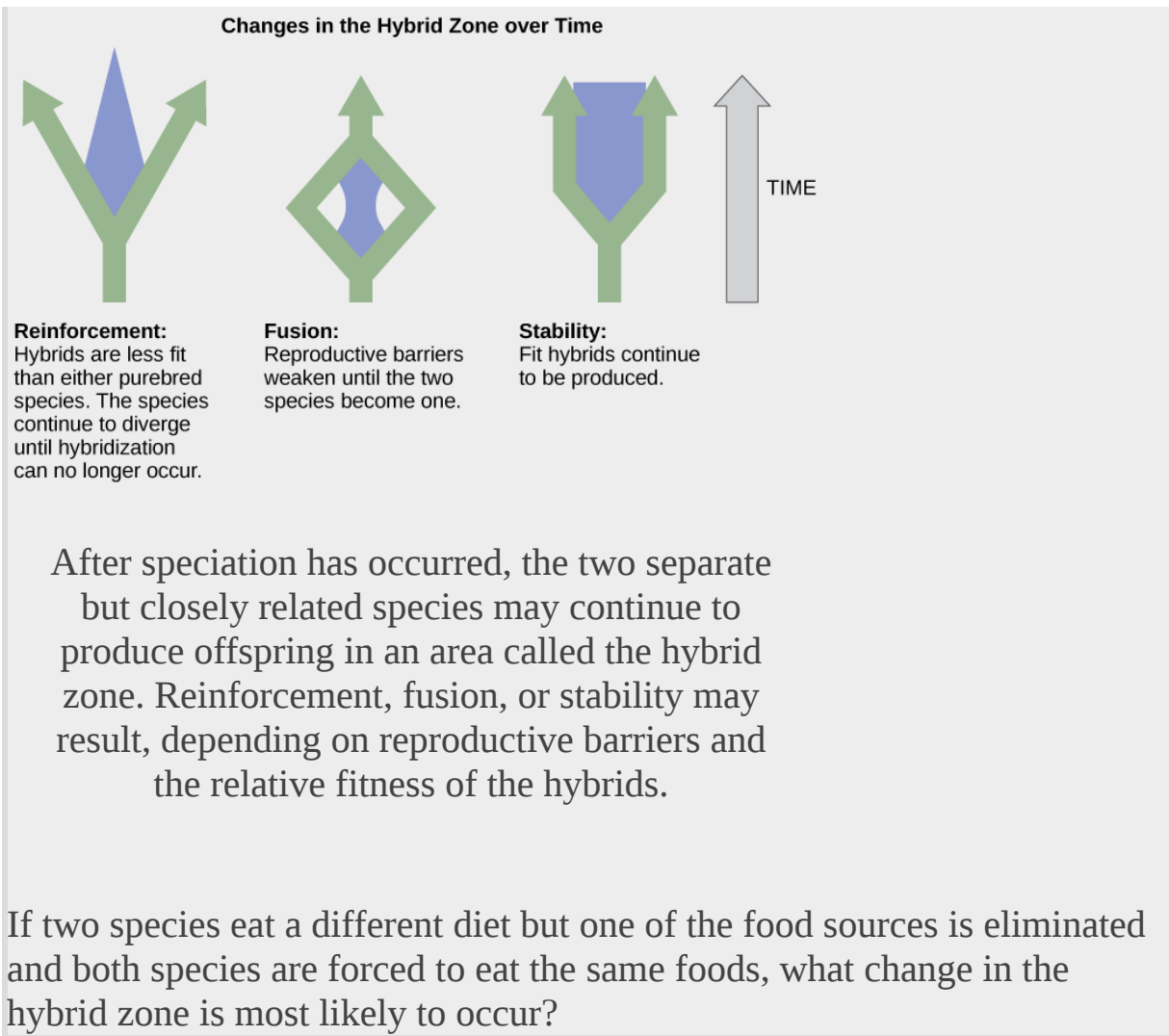
Speciation occurs over a span of evolutionary time, so when a new species arises, there is a transition period during which the closely related species continue to interact.

### Reconnection

After speciation, two species may recombine or even continue interacting indefinitely. Individual organisms will mate with any nearby individual who they are capable of breeding with. An area where two closely related species continue to interact and reproduce, forming hybrids, is called a **hybrid zone**. Over time, the hybrid zone may change depending on the fitness of the hybrids and the reproductive barriers ([link](#)). If the hybrids are less fit than the parents, reinforcement of speciation occurs, and the species continue to diverge until they can no longer mate and produce viable offspring. If reproductive barriers weaken, fusion occurs and the two species become one. Barriers remain the same if hybrids are fit and reproductive: stability may occur and hybridization continues.

#### **Note:**

Art Connection



Hybrids can be either less fit than the parents, more fit, or about the same. Usually hybrids tend to be less fit; therefore, such reproduction diminishes over time, nudging the two species to diverge further in a process called **reinforcement**. This term is used because the low success of the hybrids reinforces the original speciation. If the hybrids are as fit or more fit than the parents, the two species may fuse back into one species ([link](#)). Scientists have also observed that sometimes two species will remain separate but also continue to interact to produce some hybrid individuals; this is classified as stability because no real net change is taking place.

## Varying Rates of Speciation

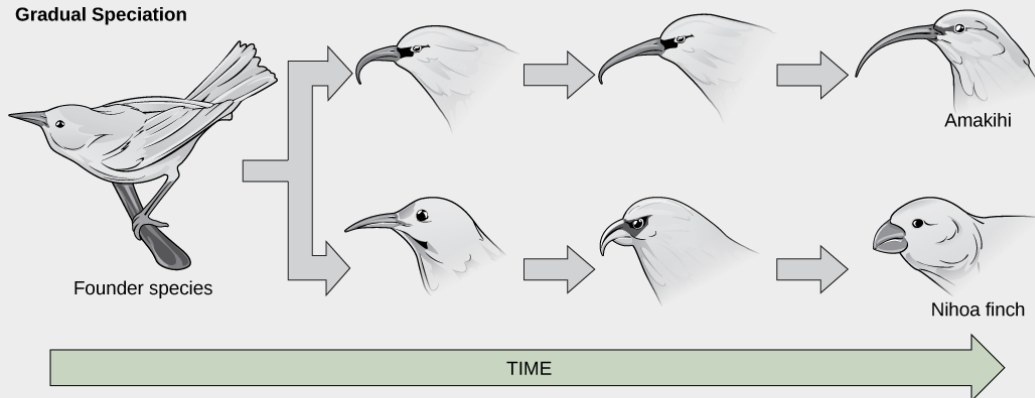
Scientists around the world study speciation, documenting observations both of living organisms and those found in the fossil record. As their ideas take shape and as research reveals new details about how life evolves, they develop models to help explain rates of speciation. In terms of how quickly speciation occurs, two patterns are currently observed: gradual speciation model and punctuated equilibrium model.

In the **gradual speciation model**, species diverge gradually over time in small steps. In the **punctuated equilibrium** model, a new species undergoes changes quickly from the parent species, and then remains largely unchanged for long periods of time afterward ([link](#)). This early change model is called punctuated equilibrium, because it begins with a punctuated or periodic change and then remains in balance afterward. While punctuated equilibrium suggests a faster tempo, it does not necessarily exclude gradualism.

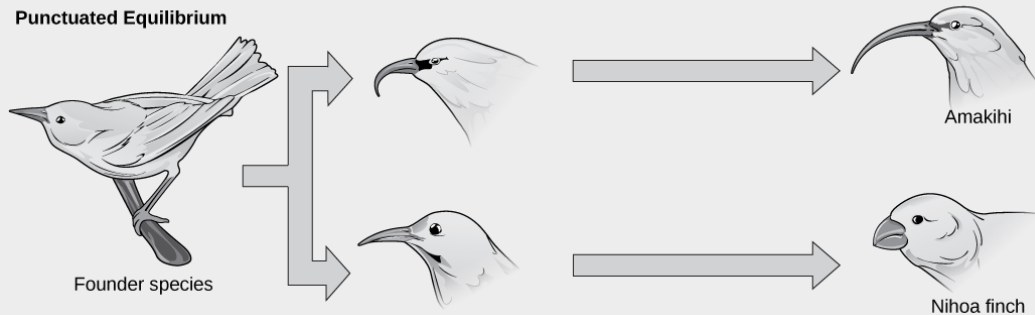
### Note:

### Art Connection

#### Gradual Speciation



#### Punctuated Equilibrium





In (a) gradual speciation, species diverge at a slow, steady pace as traits change incrementally. In (b) punctuated equilibrium, species diverge quickly and then remain unchanged for long periods of time.

Which of the following statements is false?

- a. Punctuated equilibrium is most likely to occur in a small population that experiences a rapid change in its environment.
- b. Punctuated equilibrium is most likely to occur in a large population that lives in a stable climate.
- c. Gradual speciation is most likely to occur in species that live in a stable climate.
- d. Gradual speciation and punctuated equilibrium both result in the divergence of species.

The primary influencing factor on changes in speciation rate is environmental conditions. Under some conditions, selection occurs quickly or radically. Consider a species of snails that had been living with the same basic form for many thousands of years. Layers of their fossils would appear similar for a long time. When a change in the environment takes place—such as a drop in the water level—a small number of organisms are separated from the rest in a brief period of time, essentially forming one large and one tiny population. The tiny population faces new environmental conditions. Because its gene pool quickly became so small, any variation that surfaces and that aids in surviving the new conditions becomes the predominant form.

**Note:**

[Link to Learning](#)



Visit [this website](#) to continue the speciation story of the snails.

## Section Summary

Speciation is not a precise division: overlap between closely related species can occur in areas called hybrid zones. Organisms reproduce with other similar organisms. The fitness of these hybrid offspring can affect the evolutionary path of the two species. Scientists propose two models for the rate of speciation: one model illustrates how a species can change slowly over time; the other model demonstrates how change can occur quickly from a parent generation to a new species. Both models continue to follow the patterns of natural selection.

## Art Connections

### Exercise:

#### Problem:

[\[link\]](#) If two species eat a different diet but one of the food sources is eliminated and both species are forced to eat the same foods, what change in the hybrid zone is most likely to occur?

---

#### Solution:

[\[link\]](#) Fusion is most likely to occur because the two species will interact more and similar traits in food acquisition will be selected.

### Exercise:

**Problem:** [\[link\]](#) Which of the following statements is false?

- a. Punctuated equilibrium is most likely to occur in a small population that experiences a rapid change in its environment.
- b. Punctuated equilibrium is most likely to occur in a large population that lives in a stable climate.
- c. Gradual speciation is most likely to occur in species that live in a stable climate.
- d. Gradual speciation and punctuated equilibrium both result in the evolution of new species.

---

**Solution:**

[\[link\]](#) Answer B

## Review Questions

**Exercise:**

**Problem:**

Which term is used to describe the continued divergence of species based on the low fitness of hybrid offspring?

- a. reinforcement
- b. fusion
- c. stability
- d. punctuated equilibrium

---

**Solution:**

A

**Exercise:**

**Problem:**

Which components of speciation would be least likely to be a part of punctuated equilibrium?

- a. a division of populations
- b. a change in environmental conditions
- c. ongoing gene flow among all individuals
- d. a large number of mutations taking place at once

---

**Solution:**

C

**Free Response****Exercise:**

**Problem:** What do both rate of speciation models have in common?

---

**Solution:**

Both models continue to conform to the rules of natural selection, and the influences of gene flow, genetic drift, and mutation.

**Exercise:****Problem:**

Describe a situation where hybrid reproduction would cause two species to fuse into one.

---

**Solution:**

If the hybrid offspring are as fit or more fit than the parents, reproduction would likely continue between both species and the

hybrids, eventually bringing all organisms under the umbrella of one species.

## **Glossary**

gradual speciation model

model that shows how species diverge gradually over time in small steps

hybrid zone

area where two closely related species continue to interact and reproduce, forming hybrids

punctuated equilibrium

model for rapid speciation that can occur when an event causes a small portion of a population to be cut off from the rest of the population

reinforcement

continued speciation divergence between two related species due to low fitness of hybrids between them

## Introduction

class="introduction"

Living things may be single-celled or complex, multicellular organisms. They may be plants, animals, fungi, bacteria, or archaea. This diversity results from evolution.

(credit "wolf": modification of work by Gary Kramer; credit "coral": modification of work by William Harrigan, NOAA; credit "river": modification of work by Vojtěch

Dostál;  
credit "fish"  
modification  
of work by  
Christian  
Mehlführer;  
credit  
"mushroom"  
:  
modification  
of work by  
Cory  
Zanker;  
credit "tree":  
modification  
of work by  
Joseph  
Kranak;  
credit "bee":  
modification  
of work by  
Cory  
Zanker)



All life on Earth is related. Evolutionary theory states that humans, beetles, plants, and bacteria all share a common ancestor, but that millions of years of evolution have shaped each of these organisms into the forms seen today. Scientists consider evolution a key concept to understanding life. Natural selection is one of the most dominant evolutionary forces. Natural selection acts to promote traits and behaviors that increase an organism's chances of survival and reproduction, while eliminating those traits and behaviors that are to the organism's detriment. But natural selection can only, as its name implies, select—it cannot create. The introduction of novel traits and behaviors falls on the shoulders of another evolutionary force—mutation. Mutation and other sources of variation among individuals, as well as the evolutionary forces that act upon them, alter populations and species. This combination of processes has led to the world of life we see today.



## Population Evolution

By the end of this section, you will be able to:

- Define population genetics and describe how population genetics is used in the study of the evolution of populations
- Define the Hardy-Weinberg principle and discuss its importance

The mechanisms of inheritance, or genetics, were not understood at the time Charles Darwin and Alfred Russel Wallace were developing their idea of natural selection. This lack of understanding was a stumbling block to understanding many aspects of evolution. In fact, the predominant (and incorrect) genetic theory of the time, blending inheritance, made it difficult to understand how natural selection might operate. Darwin and Wallace were unaware of the genetics work by Austrian monk Gregor Mendel, which was published in 1866, not long after publication of Darwin's book, *On the Origin of Species*. Mendel's work was rediscovered in the early twentieth century at which time geneticists were rapidly coming to an understanding of the basics of inheritance. Initially, the newly discovered particulate nature of genes made it difficult for biologists to understand how gradual evolution could occur. But over the next few decades genetics and evolution were integrated in what became known as the **modern synthesis**—the coherent understanding of the relationship between natural selection and genetics that took shape by the 1940s and is generally accepted today. In sum, the modern synthesis describes how evolutionary processes, such as natural selection, can affect a population's genetic makeup, and, in turn, how this can result in the gradual evolution of populations and species. The theory also connects this change of a population over time, called **microevolution**, with the processes that gave rise to new species and higher taxonomic groups with widely divergent characters, called **macroevolution**.

### Note:

#### Everyday Connection

#### Evolution and Flu Vaccines

Every fall, the media starts reporting on flu vaccinations and potential outbreaks. Scientists, health experts, and institutions determine

recommendations for different parts of the population, predict optimal production and inoculation schedules, create vaccines, and set up clinics to provide inoculations. You may think of the annual flu shot as a lot of media hype, an important health protection, or just a briefly uncomfortable prick in your arm. But do you think of it in terms of evolution?

The media hype of annual flu shots is scientifically grounded in our understanding of evolution. Each year, scientists across the globe strive to predict the flu strains that they anticipate being most widespread and harmful in the coming year. This knowledge is based in how flu strains have evolved over time and over the past few flu seasons. Scientists then work to create the most effective vaccine to combat those selected strains. Hundreds of millions of doses are produced in a short period in order to provide vaccinations to key populations at the optimal time.

Because viruses, like the flu, evolve very quickly (especially in evolutionary time), this poses quite a challenge. Viruses mutate and replicate at a fast rate, so the vaccine developed to protect against last year's flu strain may not provide the protection needed against the coming year's strain. Evolution of these viruses means continued adaptations to ensure survival, including adaptations to survive previous vaccines.

## Population Genetics

Recall that a gene for a particular character may have several alleles, or variants, that code for different traits associated with that character. For example, in the ABO blood type system in humans, three alleles determine the particular blood-type carbohydrate on the surface of red blood cells. Each individual in a population of diploid organisms can only carry two alleles for a particular gene, but more than two may be present in the individuals that make up the population. Mendel followed alleles as they were inherited from parent to offspring. In the early twentieth century, biologists in a field of study known as **population genetics** began to study how selective forces change a population through changes in allele and genotypic frequencies.

The **allele frequency** (or gene frequency) is the rate at which a specific allele appears within a population. Until now we have discussed evolution as a change in the characteristics of a population of organisms, but behind that phenotypic change is genetic change. In population genetics, the term evolution is defined as a change in the frequency of an allele in a population. Using the ABO blood type system as an example, the frequency of one of the alleles,  $I^A$ , is the number of copies of that allele divided by all the copies of the ABO gene in the population. For example, a study in Jordan<sup>[footnote]</sup> found a frequency of  $I^A$  to be 26.1 percent. The  $I^B$  and  $I^O$  alleles made up 13.4 percent and 60.5 percent of the alleles respectively, and all of the frequencies added up to 100 percent. A change in this frequency over time would constitute evolution in the population. Sahar S. Hanania, Dhia S. Hassawi, and Nidal M. Irshaid, “Allele Frequency and Molecular Genotypes of ABO Blood Group System in a Jordanian Population,” *Journal of Medical Sciences* 7 (2007): 51-58, doi:10.3923/jms.2007.51.58.

The allele frequency within a given population can change depending on environmental factors; therefore, certain alleles become more widespread than others during the process of natural selection. Natural selection can alter the population’s genetic makeup; for example, if a given allele confers a phenotype that allows an individual to better survive or have more offspring. Because many of those offspring will also carry the beneficial allele, and often the corresponding phenotype, they will have more offspring of their own that also carry the allele, thus, perpetuating the cycle. Over time, the allele will spread throughout the population. Some alleles will quickly become fixed in this way, meaning that every individual of the population will carry the allele, while detrimental mutations may be swiftly eliminated if derived from a dominant allele from the gene pool. The **gene pool** is the sum of all the alleles in a population.

Sometimes, allele frequencies within a population change randomly with no advantage to the population over existing allele frequencies. This phenomenon is called genetic drift. Natural selection and genetic drift usually occur simultaneously in populations and are not isolated events. It is hard to determine which process dominates because it is often nearly impossible to determine the cause of change in allele frequencies at each

occurrence. An event that initiates an allele frequency change in an isolated part of the population, which is not typical of the original population, is called the **founder effect**. Natural selection, random drift, and founder effects can lead to significant changes in the genome of a population.

## Hardy-Weinberg Principle of Equilibrium

In the early twentieth century, English mathematician Godfrey Hardy and German physician Wilhelm Weinberg stated the principle of equilibrium to describe the genetic makeup of a population. The theory, which later became known as the Hardy-Weinberg principle of equilibrium, states that a population's allele and genotype frequencies are inherently stable— unless some kind of evolutionary force is acting upon the population, neither the allele nor the genotypic frequencies would change. The Hardy-Weinberg principle assumes conditions with no mutations, migration, emigration, or selective pressure for or against genotype, plus an infinite population; while no population can satisfy those conditions, the principle offers a useful model against which to compare real population changes.

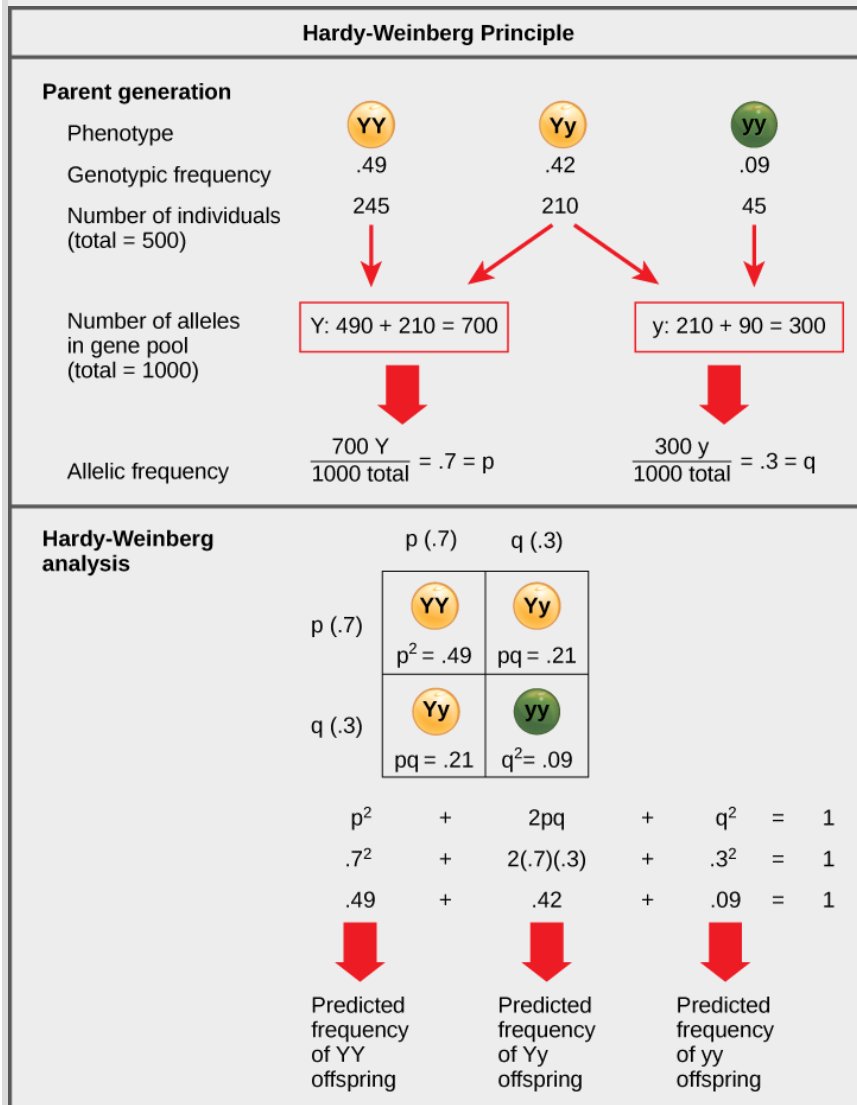
Working under this theory, population geneticists represent different alleles as different variables in their mathematical models. The variable  $p$ , for example, often represents the frequency of a particular allele, say  $Y$  for the trait of yellow in Mendel's peas, while the variable  $q$  represents the frequency of  $y$  alleles that confer the color green. If these are the only two possible alleles for a given locus in the population,  $p + q = 1$ . In other words, all the  $p$  alleles and all the  $q$  alleles make up all of the alleles for that locus that are found in the population.

But what ultimately interests most biologists is not the frequencies of different alleles, but the frequencies of the resulting genotypes, known as the population's **genetic structure**, from which scientists can surmise the distribution of phenotypes. If the phenotype is observed, only the genotype of the homozygous recessive alleles can be known; the calculations provide an estimate of the remaining genotypes. Since each individual carries two alleles per gene, if the allele frequencies ( $p$  and  $q$ ) are known, predicting the frequencies of these genotypes is a simple mathematical calculation to determine the probability of getting these genotypes if two alleles are drawn

at random from the gene pool. So in the above scenario, an individual pea plant could be pp (YY), and thus produce yellow peas; pq (Yy), also yellow; or qq (yy), and thus producing green peas ([\[link\]](#)). In other words, the frequency of pp individuals is simply  $p^2$ ; the frequency of pq individuals is  $2pq$ ; and the frequency of qq individuals is  $q^2$ . And, again, if p and q are the only two possible alleles for a given trait in the population, these genotypes frequencies will sum to one:  $p^2 + 2pq + q^2 = 1$ .

### Note:

### Art Connection



When populations are in the Hardy-Weinberg equilibrium, the allelic frequency is stable from generation to generation and the distribution of alleles can be determined from the Hardy-Weinberg equation. If the allelic frequency measured in the field differs from the predicted value, scientists can make inferences about what evolutionary forces are at play.

In plants, violet flower color (V) is dominant over white (v). If  $p = 0.8$  and  $q = 0.2$  in a population of 500 plants, how many individuals would you expect to be homozygous dominant (VV), heterozygous (Vv), and homozygous recessive (vv)? How many plants would you expect to have violet flowers, and how many would have white flowers?

In theory, if a population is at equilibrium—that is, there are no evolutionary forces acting upon it—generation after generation would have the same gene pool and genetic structure, and these equations would all hold true all of the time. Of course, even Hardy and Weinberg recognized that no natural population is immune to evolution. Populations in nature are constantly changing in genetic makeup due to drift, mutation, possibly migration, and selection. As a result, the only way to determine the exact distribution of phenotypes in a population is to go out and count them. But the Hardy-Weinberg principle gives scientists a mathematical baseline of a non-evolving population to which they can compare evolving populations and thereby infer what evolutionary forces might be at play. If the frequencies of alleles or genotypes deviate from the value expected from the Hardy-Weinberg equation, then the population is evolving.

**Note:**

[Link to Learning](#)



Use this [online calculator](#) to determine the genetic structure of a population.

## Section Summary

The modern synthesis of evolutionary theory grew out of the cohesion of Darwin's, Wallace's, and Mendel's thoughts on evolution and heredity, along with the more modern study of population genetics. It describes the evolution of populations and species, from small-scale changes among individuals to large-scale changes over paleontological time periods. To understand how organisms evolve, scientists can track populations' allele frequencies over time. If they differ from generation to generation, scientists can conclude that the population is not in Hardy-Weinberg equilibrium, and is thus evolving.

## Art Connections

### Exercise:

#### Problem:

[\[link\]](#) In plants, violet flower color (V) is dominant over white (v). If  $p = .8$  and  $q = 0.2$  in a population of 500 plants, how many individuals would you expect to be homozygous dominant (VV), heterozygous (Vv), and homozygous recessive (vv)? How many plants would you expect to have violet flowers, and how many would have white flowers?

---

#### Solution:

[\[link\]](#) The expected distribution is 320 VV, 160Vv, and 20 vv plants. Plants with VV or Vv genotypes would have violet flowers, and plants with the vv genotype would have white flowers, so a total of 480 plants would be expected to have violet flowers, and 20 plants would have white flowers.

## Review Questions

### Exercise:

**Problem:** What is the difference between micro- and macroevolution?

- a. Microevolution describes the evolution of small organisms, such as insects, while macroevolution describes the evolution of large organisms, like people and elephants.
- b. Microevolution describes the evolution of microscopic entities, such as molecules and proteins, while macroevolution describes the evolution of whole organisms.
- c. Microevolution describes the evolution of organisms in populations, while macroevolution describes the evolution of species over long periods of time.
- d. Microevolution describes the evolution of organisms over their lifetimes, while macroevolution describes the evolution of organisms over multiple generations.

---

### Solution:

C

### Exercise:

**Problem:** Population genetics is the study of:

- a. how selective forces change the allele frequencies in a population over time
- b. the genetic basis of population-wide traits



- c. whether traits have a genetic basis
  - d. the degree of inbreeding in a population
- 

**Solution:**

A

**Exercise:**

**Problem:**

Which of the following populations is not in Hardy-Weinberg equilibrium?

- a. a population with 12 homozygous recessive individuals (yy), 8 homozygous dominant individuals (YY), and 4 heterozygous individuals (Yy)
  - b. a population in which the allele frequencies do not change over time
  - c.  $p^2 + 2pq + q^2 = 1$
  - d. a population undergoing natural selection
- 

**Solution:**

D

**Exercise:**

**Problem:**

One of the original Amish colonies rose from a ship of colonists that came from Europe. The ship's captain, who had polydactyly, a rare dominant trait, was one of the original colonists. Today, we see a much higher frequency of polydactyly in the Amish population. This is an example of:

- a. natural selection
- b. genetic drift

- c. founder effect
- d. b and c

---

**Solution:**

D

**Free Response**

**Exercise:**

**Problem:**

Solve for the genetic structure of a population with 12 homozygous recessive individuals (yy), 8 homozygous dominant individuals (YY), and 4 heterozygous individuals (Yy).

---

**Solution:**

$$p = (8 \cdot 2 + 4) / 48 = .42; q = (12 \cdot 2 + 4) / 48 = .58; p^2 = .17; 2pq = .48; q^2 = .34$$

**Exercise:**

**Problem:** Explain the Hardy-Weinberg principle of equilibrium theory.

---

**Solution:**

The Hardy-Weinberg principle of equilibrium is used to describe the genetic makeup of a population. The theory states that a population's allele and genotype frequencies are inherently stable: unless some kind of evolutionary force is acting upon the population, generation after generation of the population would carry the same genes, and individuals would, as a whole, look essentially the same.

**Exercise:**

**Problem:**

Imagine you are trying to test whether a population of flowers is undergoing evolution. You suspect there is selection pressure on the color of the flower: bees seem to cluster around the red flowers more often than the blue flowers. In a separate experiment, you discover blue flower color is dominant to red flower color. In a field, you count 600 blue flowers and 200 red flowers. What would you expect the genetic structure of the flowers to be?

---

**Solution:**

Red is recessive so  $q^2 = 200/800 = 0.25$ ;  $q = 0.5$ ;  $p = 1 - q = 0.5$ ;  $p^2 = 0.25$ ;  $2pq = 0.5$ . You would expect 200 homozygous blue flowers, 400 heterozygous blue flowers, and 200 red flowers.

**Glossary**

allele frequency

(also, gene frequency) rate at which a specific allele appears within a population

founder effect

event that initiates an allele frequency change in part of the population, which is not typical of the original population

gene pool

all of the alleles carried by all of the individuals in the population

genetic structure

distribution of the different possible genotypes in a population

macroevolution

broader scale evolutionary changes seen over paleontological time

microevolution

changes in a population's genetic structure

modern synthesis

overarching evolutionary paradigm that took shape by the 1940s and is generally accepted today

population genetics

study of how selective forces change the allele frequencies in a population over time

## Population Genetics

By the end of this section, you will be able to:

- Describe the different types of variation in a population
- Explain why only heritable variation can be acted upon by natural selection
- Describe genetic drift and the bottleneck effect
- Explain how each evolutionary force can influence the allele frequencies of a population

Individuals of a population often display different phenotypes, or express different alleles of a particular gene, referred to as polymorphisms.

Populations with two or more variations of particular characteristics are called polymorphic. The distribution of phenotypes among individuals, known as the **population variation**, is influenced by a number of factors, including the population's genetic structure and the environment ([\[link\]](#)).

Understanding the sources of a phenotypic variation in a population is important for determining how a population will evolve in response to different evolutionary pressures.



The distribution of phenotypes in this litter of kittens illustrates population variation. (credit: Pieter Lanser)

## Genetic Variance

Natural selection and some of the other evolutionary forces can only act on heritable traits, namely an organism's genetic code. Because alleles are passed from parent to offspring, those that confer beneficial traits or behaviors may be selected for, while deleterious alleles may be selected against. Acquired traits, for the most part, are not heritable. For example, if an athlete works out in the gym every day, building up muscle strength, the athlete's offspring will not necessarily grow up to be a body builder. If there is a genetic basis for the ability to run fast, on the other hand, this may be passed to a child.

**Note:**

Link to Learning



Before Darwinian evolution became the prevailing theory of the field, French naturalist Jean-Baptiste Lamarck theorized that acquired traits could, in fact, be inherited; while this hypothesis has largely been unsupported, scientists have recently begun to realize that Lamarck was not completely wrong. Visit this [site](#) to learn more.

**Heritability** is the fraction of phenotype variation that can be attributed to genetic differences, or genetic variance, among individuals in a population. The greater the heritability of a population's phenotypic variation, the more susceptible it is to the evolutionary forces that act on heritable variation.

The diversity of alleles and genotypes within a population is called **genetic variance**. When scientists are involved in the breeding of a species, such as with animals in zoos and nature preserves, they try to increase a

population's genetic variance to preserve as much of the phenotypic diversity as they can. This also helps reduce the risks associated with **inbreeding**, the mating of closely related individuals, which can have the undesirable effect of bringing together deleterious recessive mutations that can cause abnormalities and susceptibility to disease. For example, a disease that is caused by a rare, recessive allele might exist in a population, but it will only manifest itself when an individual carries two copies of the allele. Because the allele is rare in a normal, healthy population with unrestricted habitat, the chance that two carriers will mate is low, and even then, only 25 percent of their offspring will inherit the disease allele from both parents. While it is likely to happen at some point, it will not happen frequently enough for natural selection to be able to swiftly eliminate the allele from the population, and as a result, the allele will be maintained at low levels in the gene pool. However, if a family of carriers begins to interbreed with each other, this will dramatically increase the likelihood of two carriers mating and eventually producing diseased offspring, a phenomenon known as **inbreeding depression**.

Changes in allele frequencies that are identified in a population can shed light on how it is evolving. In addition to natural selection, there are other evolutionary forces that could be in play: genetic drift, gene flow, mutation, nonrandom mating, and environmental variances.

## **Genetic Drift**

The theory of natural selection stems from the observation that some individuals in a population are more likely to survive longer and have more offspring than others; thus, they will pass on more of their genes to the next generation. A big, powerful male gorilla, for example, is much more likely than a smaller, weaker one to become the population's silverback, the pack's leader who mates far more than the other males of the group. The pack leader will father more offspring, who share half of his genes, and are likely to also grow bigger and stronger like their father. Over time, the genes for bigger size will increase in frequency in the population, and the population will, as a result, grow larger on average. That is, this would occur if this particular **selection pressure**, or driving selective force, were

the only one acting on the population. In other examples, better camouflage or a stronger resistance to drought might pose a selection pressure.

Another way a population's allele and genotype frequencies can change is **genetic drift** ([\[link\]](#)), which is simply the effect of chance. By chance, some individuals will have more offspring than others—not due to an advantage conferred by some genetically-encoded trait, but just because one male happened to be in the right place at the right time (when the receptive female walked by) or because the other one happened to be in the wrong place at the wrong time (when a fox was hunting).

**Note:**

Art Connection

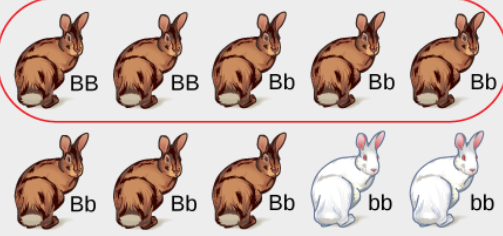


## Genetic Drift

### First generation

$p$  (B gene frequency) = .5

$q$  (b gene frequency) = .5



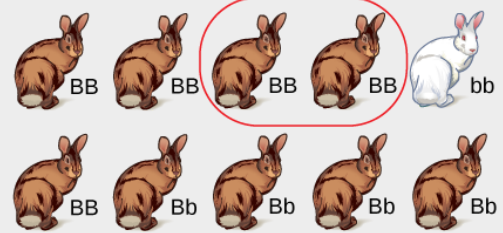
5 rabbits reproduce



### Second generation

$p$  = .7

$q$  = .3



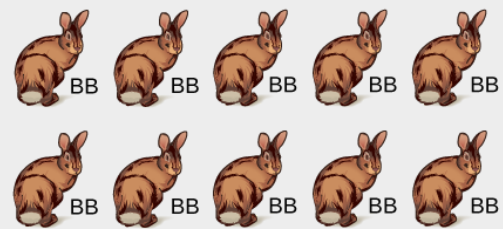
2 rabbits reproduce



### Third generation

$p$  = 1

$q$  = 0



Genetic drift in a population can lead to the elimination of an allele from a population by chance. In this example, rabbits with the brown coat color allele ( $B$ ) are dominant over rabbits with the white coat color allele ( $b$ ). In the

first generation, the two alleles occur with equal frequency in the population, resulting in  $p$  and  $q$  values of .5. Only half of the individuals reproduce, resulting in a second generation with  $p$  and  $q$  values of .7 and .3, respectively. Only two individuals in the second generation reproduce, and by chance these individuals are homozygous dominant for brown coat color. As a result, in the third generation the recessive  $b$  allele is lost.

Do you think genetic drift would happen more quickly on an island or on the mainland?

Small populations are more susceptible to the forces of genetic drift. Large populations, on the other hand, are buffered against the effects of chance. If one individual of a population of 10 individuals happens to die at a young age before it leaves any offspring to the next generation, all of its genes—1/10 of the population's gene pool—will be suddenly lost. In a population of 100, that's only 1 percent of the overall gene pool; therefore, it is much less impactful on the population's genetic structure.

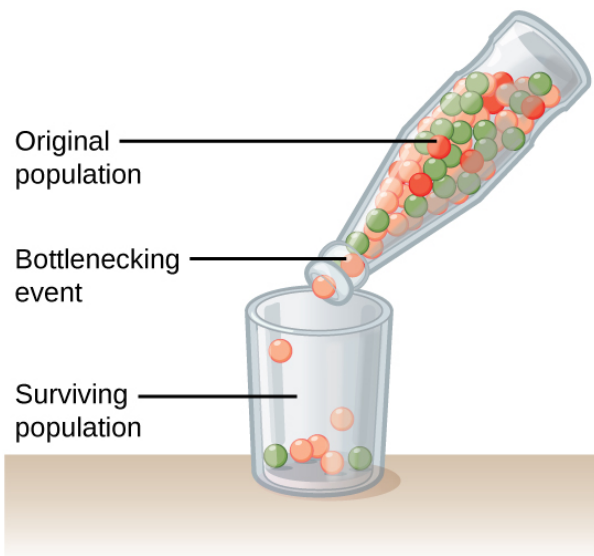
**Note:**

[Link to Learning](#)



Go to this [site](#) to watch an animation of random sampling and genetic drift in action.

Genetic drift can also be magnified by natural events, such as a natural disaster that kills—at random—a large portion of the population. Known as the **bottleneck effect**, it results in a large portion of the genome suddenly being wiped out ([link](#)). In one fell swoop, the genetic structure of the survivors becomes the genetic structure of the entire population, which may be very different from the pre-disaster population.



A chance event or catastrophe  
can reduce the genetic  
variability within a population.

Another scenario in which populations might experience a strong influence of genetic drift is if some portion of the population leaves to start a new population in a new location or if a population gets divided by a physical barrier of some kind. In this situation, those individuals are unlikely to be representative of the entire population, which results in the founder effect. The founder effect occurs when the genetic structure changes to match that of the new population's founding fathers and mothers. The founder effect is believed to have been a key factor in the genetic history of the Afrikaner population of Dutch settlers in South Africa, as evidenced by mutations that are common in Afrikaners but rare in most other populations. This is likely due to the fact that a higher-than-normal proportion of the founding colonists carried these mutations. As a result, the population expresses unusually high incidences of Huntington's disease (HD) and Fanconi anemia (FA), a genetic disorder known to cause blood marrow and congenital abnormalities—even cancer.<sup>[footnote]</sup>

A. J. Tipping et al., “Molecular and Genealogical Evidence for a Founder Effect in Fanconi Anemia Families of the Afrikaner Population of South Africa,” *PNAS* 98, no. 10 (2001): 5734-5739, doi: 10.1073/pnas.091402398.

**Note:**

Link to Learning



Watch this short video to learn more about the founder and bottleneck effects.

[https://www.openstaxcollege.org/l/founder\\_bottle](https://www.openstaxcollege.org/l/founder_bottle)

**Note:****Scientific Method Connection****Testing the Bottleneck Effect**

**Question:** How do natural disasters affect the genetic structure of a population?

**Background:** When much of a population is suddenly wiped out by an earthquake or hurricane, the individuals that survive the event are usually a random sampling of the original group. As a result, the genetic makeup of the population can change dramatically. This phenomenon is known as the bottleneck effect.

**Hypothesis:** Repeated natural disasters will yield different population genetic structures; therefore, each time this experiment is run, the results will vary.

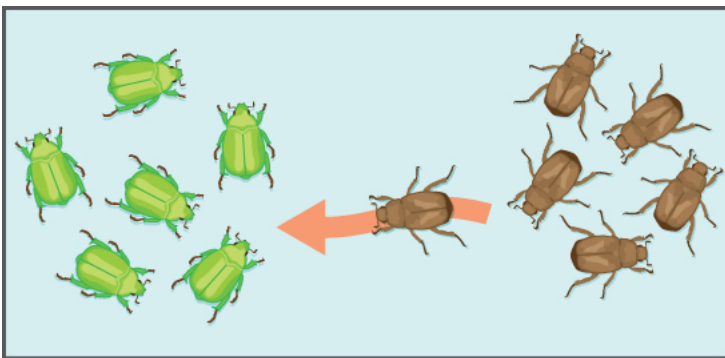
**Test the hypothesis:** Count out the original population using different colored beads. For example, red, blue, and yellow beads might represent red, blue, and yellow individuals. After recording the number of each individual in the original population, place them all in a bottle with a narrow neck that will only allow a few beads out at a time. Then, pour 1/3 of the bottle's contents into a bowl. This represents the surviving individuals after a natural disaster kills a majority of the population. Count the number of the different colored beads in the bowl, and record it. Then, place all of the beads back in the bottle and repeat the experiment four more times.

**Analyze the data:** Compare the five populations that resulted from the experiment. Do the populations all contain the same number of different colored beads, or do they vary? Remember, these populations all came from the same exact parent population.

**Form a conclusion:** Most likely, the five resulting populations will differ quite dramatically. This is because natural disasters are not selective—they kill and spare individuals at random. Now think about how this might affect a real population. What happens when a hurricane hits the Mississippi Gulf Coast? How do the seabirds that live on the beach fare?

**Gene Flow**

Another important evolutionary force is **gene flow**: the flow of alleles in and out of a population due to the migration of individuals or gametes ([link](#)). While some populations are fairly stable, others experience more flux. Many plants, for example, send their pollen far and wide, by wind or by bird, to pollinate other populations of the same species some distance away. Even a population that may initially appear to be stable, such as a pride of lions, can experience its fair share of immigration and emigration as developing males leave their mothers to seek out a new pride with genetically unrelated females. This variable flow of individuals in and out of the group not only changes the gene structure of the population, but it can also introduce new genetic variation to populations in different geological locations and habitats.



Gene flow can occur when an individual travels from one geographic location to another.

## Mutation

Mutations are changes to an organism's DNA and are an important driver of diversity in populations. Species evolve because of the accumulation of mutations that occur over time. The appearance of new mutations is the most common way to introduce novel genotypic and phenotypic variance. Some mutations are unfavorable or harmful and are quickly eliminated

from the population by natural selection. Others are beneficial and will spread through the population. Whether or not a mutation is beneficial or harmful is determined by whether it helps an organism survive to sexual maturity and reproduce. Some mutations do not do anything and can linger, unaffected by natural selection, in the genome. Some can have a dramatic effect on a gene and the resulting phenotype.

## **Nonrandom Mating**

If individuals nonrandomly mate with their peers, the result can be a changing population. There are many reasons **nonrandom mating** occurs. One reason is simple mate choice; for example, female peahens may prefer peacocks with bigger, brighter tails. Traits that lead to more matings for an individual become selected for by natural selection. One common form of mate choice, called **assortative mating**, is an individual's preference to mate with partners who are phenotypically similar to themselves.

Another cause of nonrandom mating is physical location. This is especially true in large populations spread over large geographic distances where not all individuals will have equal access to one another. Some might be miles apart through woods or over rough terrain, while others might live immediately nearby.

## **Environmental Variance**

Genes are not the only players involved in determining population variation. Phenotypes are also influenced by other factors, such as the environment ([link](#)). A beachgoer is likely to have darker skin than a city dweller, for example, due to regular exposure to the sun, an environmental factor. Some major characteristics, such as sex, are determined by the environment for some species. For example, some turtles and other reptiles have temperature-dependent sex determination (TSD). TSD means that individuals develop into males if their eggs are incubated within a certain temperature range, or females at a different temperature range.



The sex of the American alligator (*Alligator mississippiensis*) is determined by the temperature at which the eggs are incubated. Eggs incubated at 30°C produce females, and eggs incubated at 33°C produce males. (credit: Steve Hillebrand, USFWS)

Geographic separation between populations can lead to differences in the phenotypic variation between those populations. Such **geographical variation** is seen between most populations and can be significant. One type of geographic variation, called a **cline**, can be seen as populations of a given species vary gradually across an ecological gradient. Species of warm-blooded animals, for example, tend to have larger bodies in the cooler climates closer to the earth's poles, allowing them to better conserve heat. This is considered a latitudinal cline. Alternatively, flowering plants tend to bloom at different times depending on where they are along the slope of a mountain, known as an altitudinal cline.

If there is gene flow between the populations, the individuals will likely show gradual differences in phenotype along the cline. Restricted gene flow, on the other hand, can lead to abrupt differences, even speciation.



## Section Summary

Both genetic and environmental factors can cause phenotypic variation in a population. Different alleles can confer different phenotypes, and different environments can also cause individuals to look or act differently. Only those differences encoded in an individual's genes, however, can be passed to its offspring and, thus, be a target of natural selection. Natural selection works by selecting for alleles that confer beneficial traits or behaviors, while selecting against those for deleterious qualities. Genetic drift stems from the chance occurrence that some individuals in the germ line have more offspring than others. When individuals leave or join the population, allele frequencies can change as a result of gene flow. Mutations to an individual's DNA may introduce new variation into a population. Allele frequencies can also be altered when individuals do not randomly mate with others in the group.

## Art Connections

### Exercise:

#### Problem:

[\[link\]](#) Do you think genetic drift would happen more quickly on an island or on the mainland?

---

#### Solution:

[\[link\]](#) Genetic drift is likely to occur more rapidly on an island where smaller populations are expected to occur.

## Review Questions

### Exercise:

**Problem:**

When male lions reach sexual maturity, they leave their group in search of a new pride. This can alter the allele frequencies of the population through which of the following mechanisms?

- a. natural selection
- b. genetic drift
- c. gene flow
- d. random mating

---

**Solution:**

C

**Exercise:****Problem:**

Which of the following evolutionary forces can introduce new genetic variation into a population?

- a. natural selection and genetic drift
- b. mutation and gene flow
- c. natural selection and nonrandom mating
- d. mutation and genetic drift

---

**Solution:**

B

**Exercise:**

**Problem:** What is assortative mating?

- a. when individuals mate with those who are similar to themselves

- b. when individuals mate with those who are dissimilar to themselves
  - c. when individuals mate with those who are the most fit in the population
  - d. when individuals mate with those who are least fit in the population
- 

**Solution:**

A

**Exercise:**

**Problem:**

When closely related individuals mate with each other, or inbreed, the offspring are often not as fit as the offspring of two unrelated individuals. Why?

- a. Close relatives are genetically incompatible.
  - b. The DNA of close relatives reacts negatively in the offspring.
  - c. Inbreeding can bring together rare, deleterious mutations that lead to harmful phenotypes.
  - d. Inbreeding causes normally silent alleles to be expressed.
- 

**Solution:**

C

**Exercise:**

**Problem:**What is a cline?

- a. the slope of a mountain where a population lives
- b. the degree to which a mutation helps an individual survive
- c. the number of individuals in the population
- d. gradual geographic variation across an ecological gradient

---

**Solution:**

D

**Free Response****Exercise:****Problem:**

Describe a situation in which a population would undergo the bottleneck effect and explain what impact that would have on the population's gene pool.

---

**Solution:**

A hurricane kills a large percentage of a population of sand-dwelling crustaceans—only a few individuals survive. The alleles carried by those surviving individuals would represent the entire population's gene pool. If those surviving individuals are not representative of the original population, the post-hurricane gene pool will differ from the original gene pool.

**Exercise:****Problem:**

Describe natural selection and give an example of natural selection at work in a population.

---

**Solution:**

The theory of natural selection stems from the observation that some individuals in a population survive longer and have more offspring than others: thus, more of their genes are passed to the next generation. For example, a big, powerful male gorilla is much more likely than a smaller, weaker one to become the population's silverback: the pack's leader who mates far more than the other males of the group.

Therefore, the pack leader will father more offspring who share half of his genes and are likely to grow bigger and stronger like their father. Over time, the genes for bigger size will increase in frequency in the population, and the average body size, as a result, grow larger on average.

### **Exercise:**

**Problem:** Explain what a cline is and provide examples.

---

### **Solution:**

A cline is a type of geographic variation that is seen in populations of a given species that vary gradually across an ecological gradient. For example, warm-blooded animals tend to have larger bodies in the cooler climates closer to the earth's poles, allowing them to better conserve heat. This is considered a latitudinal cline. Flowering plants tend to bloom at different times depending on where they are along the slope of a mountain. This is known as an altitudinal cline.

## **Glossary**

assortative mating

when individuals tend to mate with those who are phenotypically similar to themselves

bottleneck effect

magnification of genetic drift as a result of natural events or catastrophes

cline

gradual geographic variation across an ecological gradient

gene flow

flow of alleles in and out of a population due to the migration of individuals or gametes

genetic drift

effect of chance on a population's gene pool

genetic variance

diversity of alleles and genotypes in a population

geographical variation

differences in the phenotypic variation between populations that are separated geographically

heritability

fraction of population variation that can be attributed to its genetic variance

inbreeding

mating of closely related individuals

inbreeding depression

increase in abnormalities and disease in inbreeding populations

nonrandom mating

changes in a population's gene pool due to mate choice or other forces that cause individuals to mate with certain phenotypes more than others

population variation

distribution of phenotypes in a population

selective pressure

environmental factor that causes one phenotype to be better than another

## Adaptive Evolution

By the end of this section, you will be able to:

- Explain the different ways natural selection can shape populations
- Describe how these different forces can lead to different outcomes in terms of the population variation

Natural selection only acts on the population's heritable traits: selecting for beneficial alleles and thus increasing their frequency in the population, while selecting against deleterious alleles and thereby decreasing their frequency—a process known as **adaptive evolution**. Natural selection does not act on individual alleles, however, but on entire organisms. An individual may carry a very beneficial genotype with a resulting phenotype that, for example, increases the ability to reproduce (fecundity), but if that same individual also carries an allele that results in a fatal childhood disease, that fecundity phenotype will not be passed on to the next generation because the individual will not live to reach reproductive age. Natural selection acts at the level of the individual; it selects for individuals with greater contributions to the gene pool of the next generation, known as an organism's **evolutionary (Darwinian) fitness**.

Fitness is often quantifiable and is measured by scientists in the field. However, it is not the absolute fitness of an individual that counts, but rather how it compares to the other organisms in the population. This concept, called **relative fitness**, allows researchers to determine which individuals are contributing additional offspring to the next generation, and thus, how the population might evolve.

There are several ways selection can affect population variation: stabilizing selection, directional selection, diversifying selection, frequency-dependent selection, and sexual selection. As natural selection influences the allele frequencies in a population, individuals can either become more or less genetically similar and the phenotypes displayed can become more similar or more disparate.

## Stabilizing Selection

If natural selection favors an average phenotype, selecting against extreme variation, the population will undergo **stabilizing selection** ([\[link\]](#)). In a population of mice that live in the woods, for example, natural selection is likely to favor individuals that best blend in with the forest floor and are less likely to be spotted by predators. Assuming the ground is a fairly consistent shade of brown, those mice whose fur is most closely matched to that color will be most likely to survive and reproduce, passing on their genes for their brown coat. Mice that carry alleles that make them a bit lighter or a bit darker will stand out against the ground and be more likely to fall victim to predation. As a result of this selection, the population's genetic variance will decrease.

## Directional Selection

When the environment changes, populations will often undergo **directional selection** ([\[link\]](#)), which selects for phenotypes at one end of the spectrum of existing variation. A classic example of this type of selection is the evolution of the peppered moth in eighteenth- and nineteenth-century England. Prior to the Industrial Revolution, the moths were predominately light in color, which allowed them to blend in with the light-colored trees and lichens in their environment. But as soot began spewing from factories, the trees became darkened, and the light-colored moths became easier for predatory birds to spot. Over time, the frequency of the melanic form of the moth increased because they had a higher survival rate in habitats affected by air pollution because their darker coloration blended with the sooty trees. Similarly, the hypothetical mouse population may evolve to take on a different coloration if something were to cause the forest floor where they live to change color. The result of this type of selection is a shift in the population's genetic variance toward the new, fit phenotype.

### Note:

Link to Learning





In science, sometimes things are believed to be true, and then new information comes to light that changes our understanding. The story of the peppered moth is an example: the facts behind the selection toward darker moths have recently been called into question. Read this [article](#) to learn more.

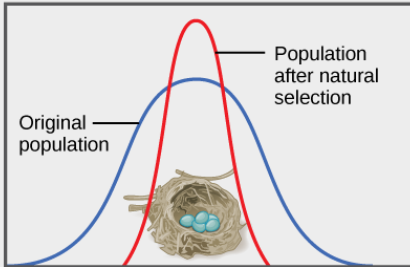
## Diversifying Selection

Sometimes two or more distinct phenotypes can each have their advantages and be selected for by natural selection, while the intermediate phenotypes are, on average, less fit. Known as **diversifying selection** ([link](#)), this is seen in many populations of animals that have multiple male forms. Large, dominant alpha males obtain mates by brute force, while small males can sneak in for furtive copulations with the females in an alpha male's territory. In this case, both the alpha males and the "sneaking" males will be selected for, but medium-sized males, which can't overtake the alpha males and are too big to sneak copulations, are selected against. Diversifying selection can also occur when environmental changes favor individuals on either end of the phenotypic spectrum. Imagine a population of mice living at the beach where there is light-colored sand interspersed with patches of tall grass. In this scenario, light-colored mice that blend in with the sand would be favored, as well as dark-colored mice that can hide in the grass. Medium-colored mice, on the other hand, would not blend in with either the grass or the sand, and would thus be more likely to be eaten by predators. The result of this type of selection is increased genetic variance as the population becomes more diverse.

## Note:

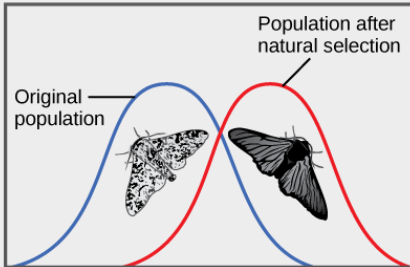
### Art Connection

(a) Stabilizing selection



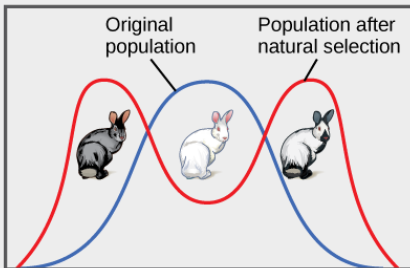
Robins typically lay four eggs, an example of stabilizing selection. Larger clutches may result in malnourished chicks, while smaller clutches may result in no viable offspring.

(b) Directional selection



Light-colored peppered moths are better camouflaged against a pristine environment; likewise, dark-colored peppered moths are better camouflaged against a sooty environment. Thus, as the Industrial Revolution progressed in nineteenth-century England, the color of the moth population shifted from light to dark, an example of directional selection.

(c) Diversifying selection



In a hypothetical population, gray and Himalayan (gray and white) rabbits are better able to blend with a rocky environment than white rabbits, resulting in diversifying selection.

Different types of natural selection can impact the distribution of phenotypes within a population. In (a) stabilizing selection, an average phenotype is favored. In (b) directional selection, a change in the environment shifts the spectrum of phenotypes observed. In (c) diversifying selection, two or more extreme phenotypes are selected for, while the average phenotype is selected against.

In recent years, factories have become cleaner, and less soot is released into the environment. What impact do you think this has had on the distribution of moth color in the population?

## Frequency-dependent Selection

Another type of selection, called **frequency-dependent selection**, favors phenotypes that are either common (positive frequency-dependent selection) or rare (negative frequency-dependent selection). An interesting example of this type of selection is seen in a unique group of lizards of the Pacific Northwest. Male common side-blotched lizards come in three throat-color patterns: orange, blue, and yellow. Each of these forms has a different reproductive strategy: orange males are the strongest and can fight other males for access to their females; blue males are medium-sized and form strong pair bonds with their mates; and yellow males ([link](#)) are the smallest, and look a bit like females, which allows them to sneak copulations. Like a game of rock-paper-scissors, orange beats blue, blue beats yellow, and yellow beats orange in the competition for females. That is, the big, strong orange males can fight off the blue males to mate with the blue's pair-bonded females, the blue males are successful at guarding their mates against yellow sneaker males, and the yellow males can sneak copulations from the potential mates of the large, polygynous orange males.



A yellow-throated side-blotched lizard is smaller than either the blue-throated or orange-throated males and appears a bit like the females of the species, allowing it to sneak copulations. (credit: “tinyfroglet”/Flickr)

In this scenario, orange males will be favored by natural selection when the population is dominated by blue males, blue males will thrive when the population is mostly yellow males, and yellow males will be selected for when orange males are the most populous. As a result, populations of side-blotched lizards cycle in the distribution of these phenotypes—in one generation, orange might be predominant, and then yellow males will begin to rise in frequency. Once yellow males make up a majority of the population, blue males will be selected for. Finally, when blue males become common, orange males will once again be favored.

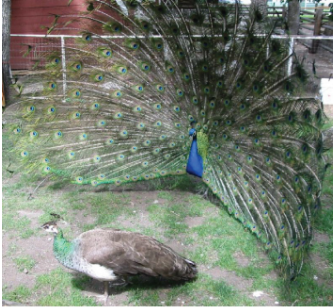
Negative frequency-dependent selection serves to increase the population’s genetic variance by selecting for rare phenotypes, whereas positive

frequency-dependent selection usually decreases genetic variance by selecting for common phenotypes.

## Sexual Selection

Males and females of certain species are often quite different from one another in ways beyond the reproductive organs. Males are often larger, for example, and display many elaborate colors and adornments, like the peacock's tail, while females tend to be smaller and duller in decoration. Such differences are known as **sexual dimorphisms** ([link](#)), which arise from the fact that in many populations, particularly animal populations, there is more variance in the reproductive success of the males than there is of the females. That is, some males—often the bigger, stronger, or more decorated males—get the vast majority of the total matings, while others receive none. This can occur because the males are better at fighting off other males, or because females will choose to mate with the bigger or more decorated males. In either case, this variation in reproductive success generates a strong selection pressure among males to get those matings, resulting in the evolution of bigger body size and elaborate ornaments to get the females' attention. Females, on the other hand, tend to get a handful of selected matings; therefore, they are more likely to select more desirable males.

Sexual dimorphism varies widely among species, of course, and some species are even sex-role reversed. In such cases, females tend to have a greater variance in their reproductive success than males and are correspondingly selected for the bigger body size and elaborate traits usually characteristic of males.



(a)



(b)



(c)

Sexual dimorphism is observed in (a) peacocks and peahens, (b) *Argiope appensa* spiders (the female spider is the large one), and in (c) wood ducks. (credit “spiders”: modification of work by “Sanba38”/Wikimedia Commons; credit “duck”: modification of work by Kevin Cole)

The selection pressures on males and females to obtain matings is known as sexual selection; it can result in the development of secondary sexual characteristics that do not benefit the individual’s likelihood of survival but help to maximize its reproductive success. Sexual selection can be so strong that it selects for traits that are actually detrimental to the individual’s survival. Think, once again, about the peacock’s tail. While it is beautiful and the male with the largest, most colorful tail is more likely to win the female, it is not the most practical appendage. In addition to being more visible to predators, it makes the males slower in their attempted escapes. There is some evidence that this risk, in fact, is why females like the big tails in the first place. The speculation is that large tails carry risk, and only the best males survive that risk: the bigger the tail, the more fit the male. This idea is known as the **handicap principle**.

The **good genes hypothesis** states that males develop these impressive ornaments to show off their efficient metabolism or their ability to fight disease. Females then choose males with the most impressive traits because it signals their genetic superiority, which they will then pass on to their offspring. Though it might be argued that females should not be picky because it will likely reduce their number of offspring, if better males father

more fit offspring, it may be beneficial. Fewer, healthier offspring may increase the chances of survival more than many, weaker offspring.

**Note:**

Link to Learning



In 1915, biologist Ronald Fisher proposed another model of sexual selection: the [Fisherian runaway model](#), which suggests that selection of certain traits is a result of sexual preference.

In both the handicap principle and the good genes hypothesis, the trait is said to be an **honest signal** of the males' quality, thus giving females a way to find the fittest mates— males that will pass the best genes to their offspring.

## No Perfect Organism

Natural selection is a driving force in evolution and can generate populations that are better adapted to survive and successfully reproduce in their environments. But natural selection cannot produce the perfect organism. Natural selection can only select on existing variation in the population; it does not create anything from scratch. Thus, it is limited by a population's existing genetic variance and whatever new alleles arise through mutation and gene flow.

Natural selection is also limited because it works at the level of individuals, not alleles, and some alleles are linked due to their physical proximity in the

genome, making them more likely to be passed on together (linkage disequilibrium). Any given individual may carry some beneficial alleles and some unfavorable alleles. It is the net effect of these alleles, or the organism's fitness, upon which natural selection can act. As a result, good alleles can be lost if they are carried by individuals that also have several overwhelmingly bad alleles; likewise, bad alleles can be kept if they are carried by individuals that have enough good alleles to result in an overall fitness benefit.

Furthermore, natural selection can be constrained by the relationships between different polymorphisms. One morph may confer a higher fitness than another, but may not increase in frequency due to the fact that going from the less beneficial to the more beneficial trait would require going through a less beneficial phenotype. Think back to the mice that live at the beach. Some are light-colored and blend in with the sand, while others are dark and blend in with the patches of grass. The dark-colored mice may be, overall, more fit than the light-colored mice, and at first glance, one might expect the light-colored mice be selected for a darker coloration. But remember that the intermediate phenotype, a medium-colored coat, is very bad for the mice—they cannot blend in with either the sand or the grass and are more likely to be eaten by predators. As a result, the light-colored mice would not be selected for a dark coloration because those individuals that began moving in that direction (began being selected for a darker coat) would be less fit than those that stayed light.

Finally, it is important to understand that not all evolution is adaptive. While natural selection selects the fittest individuals and often results in a more fit population overall, other forces of evolution, including genetic drift and gene flow, often do the opposite: introducing deleterious alleles to the population's gene pool. Evolution has no purpose—it is not changing a population into a preconceived ideal. It is simply the sum of the various forces described in this chapter and how they influence the genetic and phenotypic variance of a population.

## **Section Summary**



Because natural selection acts to increase the frequency of beneficial alleles and traits while decreasing the frequency of deleterious qualities, it is adaptive evolution. Natural selection acts at the level of the individual, selecting for those that have a higher overall fitness compared to the rest of the population. If the fit phenotypes are those that are similar, natural selection will result in stabilizing selection, and an overall decrease in the population's variation. Directional selection works to shift a population's variance toward a new, fit phenotype, as environmental conditions change. In contrast, diversifying selection results in increased genetic variance by selecting for two or more distinct phenotypes.

Other types of selection include frequency-dependent selection, in which individuals with either common (positive frequency-dependent selection) or rare (negative frequency-dependent selection) are selected for. Finally, sexual selection results from the fact that one sex has more variance in the reproductive success than the other. As a result, males and females experience different selective pressures, which can often lead to the evolution of phenotypic differences, or sexual dimorphisms, between the two.

## Art Connection

### Exercise:

#### Problem:

[\[link\]](#) In recent years, factories have become cleaner, and less soot is released into the environment. What impact do you think this has had on the distribution of moth color in the population?

---

#### Solution:

[\[link\]](#) Moths have shifted to a lighter color.

## Review Questions

### Exercise:

**Problem:**

Which type of selection results in greater genetic variance in a population?

- a. stabilizing selection
- b. directional selection
- c. diversifying selection
- d. positive frequency-dependent selection

---

**Solution:**

C

**Exercise:**

**Problem:**

When males and females of a population look or act differently, it is referred to as \_\_\_\_\_.

- a. sexual dimorphism
- b. sexual selection
- c. diversifying selection
- d. a cline

---

**Solution:**

A

**Exercise:**

**Problem:** The good genes hypothesis is a theory that explains what?

- a. why more fit individuals are more likely to have more offspring
- b. why alleles that confer beneficial traits or behaviors are selected for by natural selection
- c. why some deleterious mutations are maintained in the population

d. why individuals of one sex develop impressive ornamental traits

---

**Solution:**

D

## Free Response

**Exercise:**

**Problem:**

Give an example of a trait that may have evolved as a result of the handicap principle and explain your reasoning.

---

**Solution:**

The peacock's tail is a good example of the handicap principle. The tail, which makes the males more visible to predators and less able to escape, is clearly a disadvantage to the bird's survival. But because it is a disadvantage, only the most fit males should be able to survive with it. Thus, the tail serves as an honest signal of quality to the females of the population; therefore, the male will earn more matings and greater reproductive success.

**Exercise:**

**Problem:**

List the ways in which evolution can affect population variation and describe how they influence allele frequencies.

---

**Solution:**

There are several ways evolution can affect population variation: stabilizing selection, directional selection, diversifying selection, frequency-dependent selection, and sexual selection. As these influence the allele frequencies in a population, individuals can either

become more or less related, and the phenotypes displayed can become more similar or more disparate.

## **Glossary**

adaptive evolution

increase in frequency of beneficial alleles and decrease in deleterious alleles due to selection

directional selection

selection that favors phenotypes at one end of the spectrum of existing variation

diversifying selection

selection that favors two or more distinct phenotypes

evolutionary fitness

(also, Darwinian fitness) individual's ability to survive and reproduce

frequency-dependent selection

selection that favors phenotypes that are either common (positive frequency-dependent selection) or rare (negative frequency-dependent selection)

good genes hypothesis

theory of sexual selection that argues individuals develop impressive ornaments to show off their efficient metabolism or ability to fight disease

handicap principle

theory of sexual selection that argues only the fittest individuals can afford costly traits

honest signal

trait that gives a truthful impression of an individual's fitness

relative fitness

individual's ability to survive and reproduce relative to the rest of the population

sexual dimorphism

phenotypic difference between the males and females of a population

stabilizing selection

selection that favors average phenotypes

## Introduction

class="introduction"

The life of a  
bee is very  
different  
from the life  
of a flower,  
but the two  
organisms  
are related.

Both are  
members  
the domain  
Eukarya and  
have cells  
containing  
many  
similar  
organelles,  
genes, and  
proteins.

(credit:  
modificatio  
n of work  
by John  
Beetham)



This bee and *Echinacea* flower ([\[link\]](#)) could not look more different, yet they are related, as are all living organisms on Earth. By following pathways of similarities and changes—both visible and genetic—scientists seek to map the evolutionary past of how life developed from single-celled organisms to the tremendous collection of creatures that have germinated, crawled, floated, swam, flown, and walked on this planet.

## Organizing Life on Earth

By the end of this section, you will be able to:

- Discuss the need for a comprehensive classification system
- List the different levels of the taxonomic classification system
- Describe how systematics and taxonomy relate to phylogeny
- Discuss the components and purpose of a phylogenetic tree

In scientific terms, the evolutionary history and relationship of an organism or group of organisms is called its **phylogeny**. A phylogeny describes the relationships of an organism, such as from which organisms it is thought to have evolved, to which species it is most closely related, and so forth.

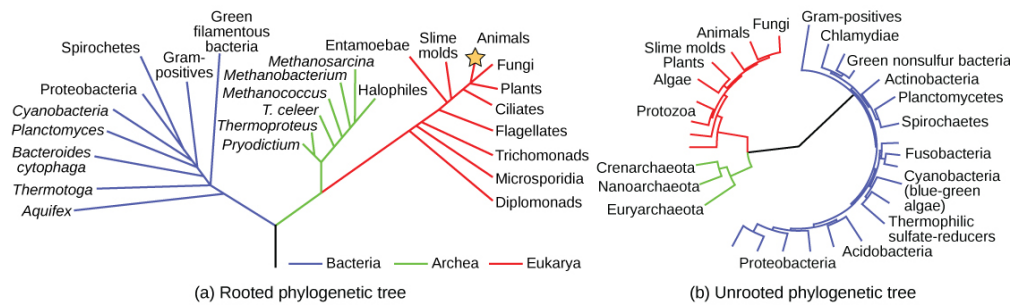
Phylogenetic relationships provide information on shared ancestry but not necessarily on how organisms are similar or different.

## Phylogenetic Trees

Scientists use a tool called a phylogenetic tree to show the evolutionary pathways and connections among organisms. A **phylogenetic tree** is a diagram used to reflect evolutionary relationships among organisms or groups of organisms. Scientists consider phylogenetic trees to be a hypothesis of the evolutionary past since one cannot go back to confirm the proposed relationships. In other words, a “tree of life” can be constructed to illustrate when different organisms evolved and to show the relationships among different organisms ([\[link\]](#)).

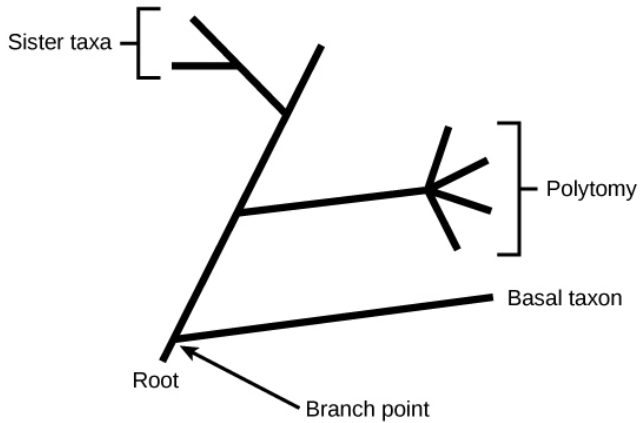
Unlike a taxonomic classification diagram, a phylogenetic tree can be read like a map of evolutionary history. Many phylogenetic trees have a single lineage at the base representing a common ancestor. Scientists call such trees **rooted**, which means there is a single ancestral lineage (typically drawn from the bottom or left) to which all organisms represented in the diagram relate. Notice in the rooted phylogenetic tree that the three domains — Bacteria, Archaea, and Eukarya—diverge from a single point and branch off. The small branch that plants and animals (including humans) occupy in this diagram shows how recent and miniscule these groups are compared with other organisms. Unrooted trees don’t show a common ancestor but do show relationships among species.





Both of these phylogenetic trees shows the relationship of the three domains of life—Bacteria, Archaea, and Eukarya—but the (a) rooted tree attempts to identify when various species diverged from a common ancestor while the (b) unrooted tree does not. (credit a: modification of work by Eric Gaba)

In a rooted tree, the branching indicates evolutionary relationships ([\[link\]](#)). The point where a split occurs, called a **branch point**, represents where a single lineage evolved into a distinct new one. A lineage that evolved early from the root and remains unbranched is called **basal taxon**. When two lineages stem from the same branch point, they are called **sister taxa**. A branch with more than two lineages is called a **polytomy** and serves to illustrate where scientists have not definitively determined all of the relationships. It is important to note that although sister taxa and polytomy do share an ancestor, it does not mean that the groups of organisms split or evolved from each other. Organisms in two taxa may have split apart at a specific branch point, but neither taxa gave rise to the other.



The root of a phylogenetic tree indicates that an ancestral lineage gave rise to all organisms on the tree. A branch point indicates where two lineages diverged. A lineage that evolved early and remains unbranched is a basal taxon. When two lineages stem from the same branch point, they are sister taxa. A branch with more than two lineages is a polytomy.

The diagrams above can serve as a pathway to understanding evolutionary history. The pathway can be traced from the origin of life to any individual species by navigating through the evolutionary branches between the two points. Also, by starting with a single species and tracing back towards the "trunk" of the tree, one can discover that species' ancestors, as well as where lineages share a common ancestry. In addition, the tree can be used to study entire groups of organisms.

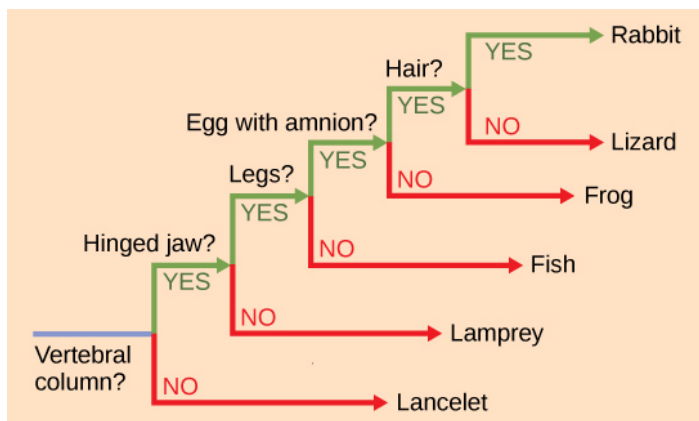
Another point to mention on phylogenetic tree structure is that rotation at branch points does not change the information. For example, if a branch point was rotated and the taxon order changed, this would not alter the

information because the evolution of each taxon from the branch point was independent of the other.

Many disciplines within the study of biology contribute to understanding how past and present life evolved over time; these disciplines together contribute to building, updating, and maintaining the “tree of life.” Information is used to organize and classify organisms based on evolutionary relationships in a scientific field called **systematics**. Data may be collected from fossils, from studying the structure of body parts or molecules used by an organism, and by DNA analysis. By combining data from many sources, scientists can put together the phylogeny of an organism; since phylogenetic trees are hypotheses, they will continue to change as new types of life are discovered and new information is learned.

## Limitations of Phylogenetic Trees

It may be easy to assume that more closely related organisms look more alike, and while this is often the case, it is not always true. If two closely related lineages evolved under significantly varied surroundings or after the evolution of a major new adaptation, it is possible for the two groups to appear more different than other groups that are not as closely related. For example, the phylogenetic tree in [\[link\]](#) shows that lizards and rabbits both have amniotic eggs, whereas frogs do not; yet lizards and frogs appear more similar than lizards and rabbits.



This ladder-like phylogenetic tree of vertebrates is rooted by an organism that lacked a vertebral column. At each branch point, organisms with different characters are placed in different groups based on the characteristics they share.

Another aspect of phylogenetic trees is that, unless otherwise indicated, the branches do not account for length of time, only the evolutionary order. In other words, the length of a branch does not typically mean more time passed, nor does a short branch mean less time passed— unless specified on the diagram. For example, in [\[link\]](#), the tree does not indicate how much time passed between the evolution of amniotic eggs and hair. What the tree does show is the order in which things took place. Again using [\[link\]](#), the tree shows that the oldest trait is the vertebral column, followed by hinged jaws, and so forth. Remember that any phylogenetic tree is a part of the greater whole, and like a real tree, it does not grow in only one direction after a new branch develops. So, for the organisms in [\[link\]](#), just because a vertebral column evolved does not mean that invertebrate evolution ceased, it only means that a new branch formed. Also, groups that are not closely related, but evolve under similar conditions, may appear more phenotypically similar to each other than to a close relative.

**Note:****Link to Learning**

Head to this [website](#) to see interactive exercises that allow you to explore the evolutionary relationships among species.

## The Levels of Classification

**Taxonomy** (which literally means “arrangement law”) is the science of classifying organisms to construct internationally shared classification systems with each organism placed into more and more inclusive groupings. Think about how a grocery store is organized. One large space is divided into departments, such as produce, dairy, and meats. Then each department further divides into aisles, then each aisle into categories and brands, and then finally a single product. This organization from larger to smaller, more specific categories is called a hierarchical system.

The taxonomic classification system (also called the Linnaean system after its inventor, Carl Linnaeus, a Swedish botanist, zoologist, and physician) uses a hierarchical model. Moving from the point of origin, the groups become more specific, until one branch ends as a single species. For example, after the common beginning of all life, scientists divide organisms into three large categories called a domain: Bacteria, Archaea, and Eukarya. Within each domain is a second category called a **kingdom**. After kingdoms, the subsequent categories of increasing specificity are: **phylum**, **class**, **order**, **family**, **genus**, and **species** ([link](#)).



**Subspecies:** *Canus lupus familiaris*

**Species:** *Canis lupus*

**Genus:** *Canis*

**Family:** Canidae

**Order:** Carnivora

**Class:** Mammalia

**Phylum:** Chordata

**Kingdom:** Animalia

**Domain:** Eukarya

The taxonomic classification system uses a hierarchical model to organize living organisms into increasingly specific categories. The common dog, *Canis lupus familiaris*, is a subspecies of *Canis lupus*, which also includes the wolf and dingo. (credit “dog”:














































modification of work by Janneke  
Vreugdenhil)

The kingdom Animalia stems from the Eukarya domain. For the common dog, the classification levels would be as shown in [\[link\]](#). Therefore, the full name of an organism technically has eight terms. For the dog, it is: Eukarya, Animalia, Chordata, Mammalia, Carnivora, Canidae, *Canis*, and *lupus*. Notice that each name is capitalized except for species, and the genus and species names are italicized. Scientists generally refer to an organism only by its genus and species, which is its two-word scientific name, in what is called **binomial nomenclature**. Therefore, the scientific name of the dog is *Canis lupus*. The name at each level is also called a **taxon**. In other words, dogs are in order Carnivora. Carnivora is the name of the taxon at the order level; Canidae is the taxon at the family level, and so forth. Organisms also have a common name that people typically use, in this case, dog. Note that the dog is additionally a subspecies: the “*familiaris*” in *Canis lupus familiaris*. Subspecies are members of the same species that are capable of mating and reproducing viable offspring, but they are considered separate subspecies due to geographic or behavioral isolation or other factors.

[\[link\]](#) shows how the levels move toward specificity with other organisms. Notice how the dog shares a domain with the widest diversity of organisms, including plants and butterflies. At each sublevel, the organisms become more similar because they are more closely related. Historically, scientists classified organisms using characteristics, but as DNA technology developed, more precise phylogenies have been determined.

**Note:**

Art Connection

<b>Subspecies:</b> <i>Canis lupus familiaris</i>	 Dog								
<b>Species:</b> <i>Canis lupus</i>	 Wolf	 Dog							
<b>Genus:</b> <i>Canis</i>	 Jackal	 Wolf	 Dog						
<b>Family:</b> Canidae	 Fox	 Jackal	 Wolf	 Dog					
<b>Order:</b> Carnivora	 Cat	 Fox	 Jackal	 Wolf	 Dog				
<b>Class:</b> Mammalia	 Rabbit	 Cat	 Fox	 Jackal	 Wolf	 Dog			
<b>Phylum:</b> Chordata	 Fish	 Rabbit	 Cat	 Fox	 Jackal	 Wolf	 Dog		
<b>Kingdom:</b> Animalia	 Insect	 Fish	 Rabbit	 Cat	 Fox	 Jackal	 Wolf	 Dog	
<b>Domain:</b> Eukarya	 Plant	 Insect	 Fish	 Rabbit	 Cat	 Fox	 Jackal	 Wolf	 Dog

At each sublevel in the taxonomic classification system, organisms become more similar. Dogs and wolves are the same species because they can breed and produce viable offspring, but they are different enough to be classified as different subspecies. (credit “plant”: modification of work by "berduchwal"/Flickr; credit “insect”: modification of work by Jon Sullivan; credit “fish”: modification of work by Christian Mehlführer; credit “rabbit”: modification of work by Aidan Wojtas; credit “cat”: modification of work by Jonathan Lidbeck; credit “fox”: modification of work by Kevin Bacher, NPS; credit “jackal”: modification of work by Thomas A. Hermann, NBII, USGS; credit “wolf”: modification



of work by Robert Dewar; credit “dog”: modification of work by "digital\_image\_fan"/Flickr)

At what levels are cats and dogs considered to be part of the same group?

**Note:**

Link to Learning



Visit this [website](#) to classify three organisms—bear, orchid, and sea cucumber—from kingdom to species. To launch the game, under Classifying Life, click the picture of the bear or the Launch Interactive button.

Recent genetic analysis and other advancements have found that some earlier phylogenetic classifications do not align with the evolutionary past; therefore, changes and updates must be made as new discoveries occur. Recall that phylogenetic trees are hypotheses and are modified as data becomes available. In addition, classification historically has focused on grouping organisms mainly by shared characteristics and does not necessarily illustrate how the various groups relate to each other from an evolutionary perspective. For example, despite the fact that a hippopotamus resembles a pig more than a whale, the hippopotamus may be the closest living relative of the whale.

## Section Summary

Scientists continually gain new information that helps understand the evolutionary history of life on Earth. Each group of organisms went through its own evolutionary journey, called its phylogeny. Each organism shares relatedness with others, and based on morphologic and genetic evidence, scientists attempt to map the evolutionary pathways of all life on Earth. Historically, organisms were organized into a taxonomic classification system. However, today many scientists build phylogenetic trees to illustrate evolutionary relationships.

## Art Connections

### Exercise:

#### Problem:

[\[link\]](#) At what levels are cats and dogs considered to be part of the same group?

---

#### Solution:

[\[link\]](#) Cats and dogs are part of the same group at five levels: both are in the domain Eukarya, the kingdom Animalia, the phylum Chordata, the class Mammalia, and the order Carnivora.

## Review Questions

### Exercise:

**Problem:** What is used to determine phylogeny?

- a. mutations
  - b. DNA
  - c. evolutionary history
  - d. organisms on earth
- 

#### Solution:

C

**Exercise:**

**Problem:** What do scientists in the field of systematics accomplish?

- a. discover new fossil sites
  - b. organize and classify organisms
  - c. name new species
  - d. communicate among field biologists
- 

**Solution:**

B

**Exercise:**

**Problem:**

Which statement about the taxonomic classification system is correct?

- a. There are more domains than kingdoms.
  - b. Kingdoms are the top category of classification.
  - c. Classes are divisions of orders.
  - d. Subspecies are the most specific category of classification.
- 

**Solution:**

D

**Exercise:**

**Problem:**

On a phylogenetic tree, which term refers to lineages that diverged from the same place?

- a. sister taxa
- b. basal taxa

- c. rooted taxa
  - d. dichotomous taxa
- 

**Solution:**

A

## Free Response

**Exercise:**

**Problem:** How does a phylogenetic tree relate to the passing of time?

---

**Solution:**

The phylogenetic tree shows the order in which evolutionary events took place and in what order certain characteristics and organisms evolved in relation to others. It does not relate to time.

**Exercise:**

**Problem:**

Some organisms that appear very closely related on a phylogenetic tree may not actually be closely related. Why is this?

---

**Solution:**

In most cases, organisms that appear closely related actually are; however, there are cases where organisms evolved through convergence and appear closely related but are not.

**Exercise:**

**Problem:**

List the different levels of the taxonomic classification system.

---

**Solution:**

domain, kingdom, phylum, class, order, family, genus, species

**Glossary**

basal taxon

branch on a phylogenetic tree that has not diverged significantly from the root ancestor

binomial nomenclature

system of two-part scientific names for an organism, which includes genus and species names

branch point

node on a phylogenetic tree where a single lineage splits into distinct new ones

class

division of phylum in the taxonomic classification system

family

division of order in the taxonomic classification system

genus

division of family in the taxonomic classification system; the first part of the binomial scientific name

kingdom

division of domain in the taxonomic classification system

order

division of class in the taxonomic classification system

phylogenetic tree

diagram used to reflect the evolutionary relationships among organisms or groups of organisms

phylogeny

evolutionary history and relationship of an organism or group of organisms

phylum

(plural: phyla) division of kingdom in the taxonomic classification system

polytomy

branch on a phylogenetic tree with more than two groups or taxa

rooted

single ancestral lineage on a phylogenetic tree to which all organisms represented in the diagram relate

sister taxa

two lineages that diverged from the same branch point

systematics

field of organizing and classifying organisms based on evolutionary relationships

taxon

(plural: taxa) single level in the taxonomic classification system

taxonomy

science of classifying organisms

## Determining Evolutionary Relationships

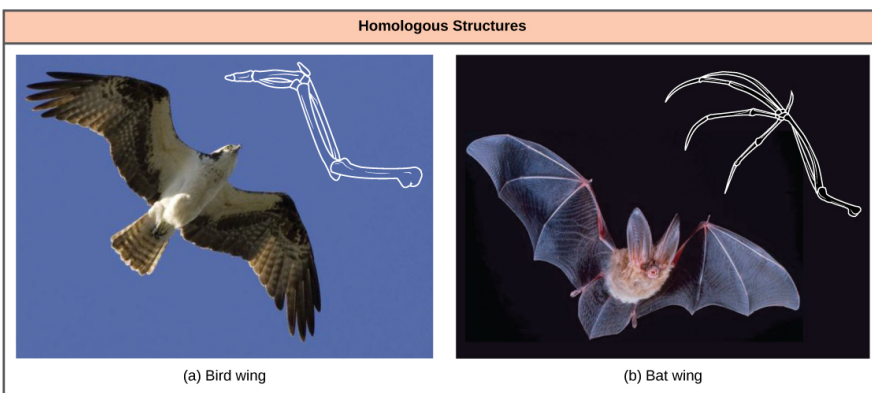
By the end of this section, you will be able to:

- Compare homologous and analogous traits
- Discuss the purpose of cladistics
- Describe maximum parsimony

Scientists must collect accurate information that allows them to make evolutionary connections among organisms. Similar to detective work, scientists must use evidence to uncover the facts. In the case of phylogeny, evolutionary investigations focus on two types of evidence: morphologic (form and function) and genetic.

## Two Options for Similarities

In general, organisms that share similar physical features and genomes tend to be more closely related than those that do not. Such features that overlap both morphologically (in form) and genetically are referred to as homologous structures; they stem from developmental similarities that are based on evolution. For example, the bones in the wings of bats and birds have homologous structures ([link](#)).



Bat and bird wings are homologous structures, indicating that bats and birds share a common evolutionary past. (credit a: modification of work

by Steve Hillebrand, USFWS; credit b:  
modification of work by U.S. DOI BLM)

Notice it is not simply a single bone, but rather a grouping of several bones arranged in a similar way. The more complex the feature, the more likely any kind of overlap is due to a common evolutionary past. Imagine two people from different countries both inventing a car with all the same parts and in exactly the same arrangement without any previous or shared knowledge. That outcome would be highly improbable. However, if two people both invented a hammer, it would be reasonable to conclude that both could have the original idea without the help of the other. The same relationship between complexity and shared evolutionary history is true for homologous structures in organisms.

## Misleading Appearances

Some organisms may be very closely related, even though a minor genetic change caused a major morphological difference to make them look quite different. Similarly, unrelated organisms may be distantly related, but appear very much alike. This usually happens because both organisms were in common adaptations that evolved within similar environmental conditions. When similar characteristics occur because of environmental constraints and not due to a close evolutionary relationship, it is called an **analogy** or homoplasy. For example, insects use wings to fly like bats and birds, but the wing structure and embryonic origin is completely different. These are called analogous structures ([\[link\]](#)).

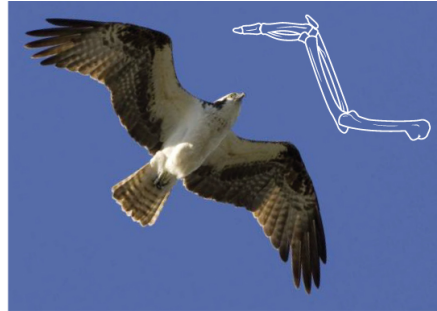
Similar traits can be either homologous or analogous. Homologous structures share a similar embryonic origin; analogous organs have a similar function. For example, the bones in the front flipper of a whale are homologous to the bones in the human arm. These structures are not analogous. The wings of a butterfly and the wings of a bird are analogous but not homologous. Some structures are both analogous and homologous: the wings of a bird and the wings of a bat are both homologous and



analogous. Scientists must determine which type of similarity a feature exhibits to decipher the phylogeny of the organisms being studied.



(a) Bat wing



(b) Bird wing



(c) Insect wing

The (c) wing of a honeybee is similar in shape to a (b) bird wing and (a) bat wing, and it serves the same function. However, the honeybee wing is not composed of bones and has a distinctly different structure and embryonic origin. These wing types (insect versus bat and bird) illustrate an analogy—similar structures that do not share an evolutionary history. (credit a: modification of work by Steve Hillebrand, USFWS; credit b: modification of work by U.S. DOI BLM; credit c: modification of work by Jon Sullivan)

**Note:**

Link to Learning



This [website](#) has several examples to show how appearances can be misleading in understanding the phylogenetic relationships of organisms.

## Molecular Comparisons

With the advancement of DNA technology, the area of **molecular systematics**, which describes the use of information on the molecular level including DNA analysis, has blossomed. New computer programs not only confirm many earlier classified organisms, but also uncover previously made errors. As with physical characteristics, even the DNA sequence can be tricky to read in some cases. For some situations, two very closely related organisms can appear unrelated if a mutation occurred that caused a shift in the genetic code. An insertion or deletion mutation would move each nucleotide base over one place, causing two similar codes to appear unrelated.

Sometimes two segments of DNA code in distantly related organisms randomly share a high percentage of bases in the same locations, causing these organisms to appear closely related when they are not. For both of these situations, computer technologies have been developed to help identify the actual relationships, and, ultimately, the coupled use of both morphologic and molecular information is more effective in determining phylogeny.

**Note:**

## Evolution Connection

### Why Does Phylogeny Matter?

Evolutionary biologists could list many reasons why understanding phylogeny is important to everyday life in human society. For botanists, phylogeny acts as a guide to discovering new plants that can be used to benefit people. Think of all the ways humans use plants—food, medicine, and clothing are a few examples. If a plant contains a compound that is effective in treating cancer, scientists might want to examine all of the relatives of that plant for other useful drugs.

A research team in China identified a segment of DNA thought to be common to some medicinal plants in the family Fabaceae (the legume family) and worked to identify which species had this segment ([link](#)). After testing plant species in this family, the team found a DNA marker (a known location on a chromosome that enabled them to identify the species) present. Then, using the DNA to uncover phylogenetic relationships, the team could identify whether a newly discovered plant was in this family and assess its potential medicinal properties.



*Dalbergia sissoo* (*D. sissoo*) is in the Fabaceae, or legume family.

Scientists found that *D. sissoo* shares a DNA marker with species within the Fabaceae family that have antifungal properties.

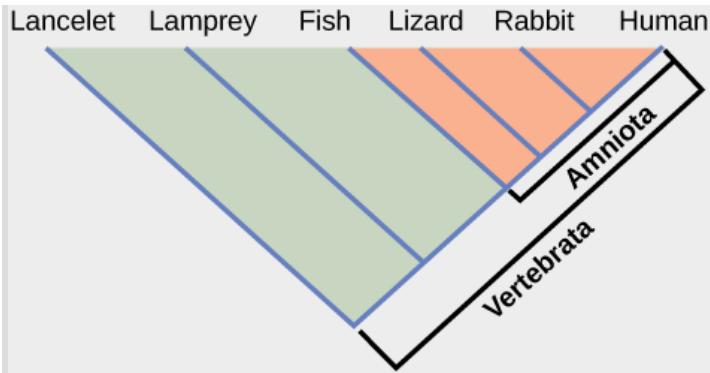
Subsequently, *D. sissoo* was shown to have fungicidal activity, supporting the idea that DNA markers can be used to screen for plants with potential medicinal properties.

## Building Phylogenetic Trees

How do scientists construct phylogenetic trees? After the homologous and analogous traits are sorted, scientists often organize the homologous traits using a system called **cladistics**. This system sorts organisms into clades: groups of organisms that descended from a single ancestor. For example, in [\[link\]](#), all of the organisms in the orange region evolved from a single ancestor that had amniotic eggs. Consequently, all of these organisms also have amniotic eggs and make a single clade, also called a **monophyletic group**. Clades must include all of the descendants from a branch point.

### Note:

Art Connection

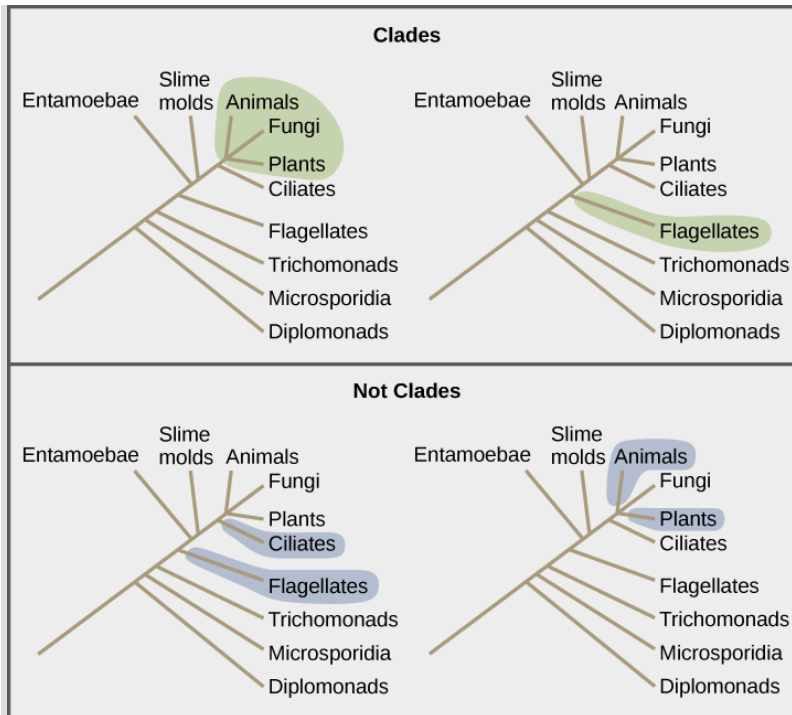


Lizards, rabbits, and humans all descend from a common ancestor that had an amniotic egg. Thus, lizards, rabbits, and humans all belong to the clade Amniota. Vertebrata is a larger clade that also includes fish and lamprey.

Which animals in this figure belong to a clade that includes animals with hair? Which evolved first, hair or the amniotic egg?

Clades can vary in size depending on which branch point is being referenced. The important factor is that all of the organisms in the clade or monophyletic group stem from a single point on the tree. This can be remembered because monophyletic breaks down into “mono,” meaning one, and “phyletic,” meaning evolutionary relationship. [\[link\]](#) shows various examples of clades. Notice how each clade comes from a single point, whereas the non-clade groups show branches that do not share a single point.

**Note:**  
Art Connection



All the organisms within a clade stem from a single point on the tree. A clade may contain multiple groups, as in the case of animals, fungi and plants, or a single group, as in the case of flagellates. Groups that diverge at a different branch point, or that do not include all groups in a single branch point, are not considered clades.

What is the largest clade in this diagram?

## Shared Characteristics

Organisms evolve from common ancestors and then diversify. Scientists use the phrase “descent with modification” because even though related organisms have many of the same characteristics and genetic codes, changes occur. This pattern repeats over and over as one goes through the phylogenetic tree of life:

1. A change in the genetic makeup of an organism leads to a new trait which becomes prevalent in the group.
2. Many organisms descend from this point and have this trait.
3. New variations continue to arise: some are adaptive and persist, leading to new traits.
4. With new traits, a new branch point is determined (go back to step 1 and repeat).

If a characteristic is found in the ancestor of a group, it is considered a **shared ancestral character** because all of the organisms in the taxon or clade have that trait. The vertebrate in [\[link\]](#) is a shared ancestral character. Now consider the amniotic egg characteristic in the same figure. Only some of the organisms in [\[link\]](#) have this trait, and to those that do, it is called a **shared derived character** because this trait derived at some point but does not include all of the ancestors in the tree.

The tricky aspect to shared ancestral and shared derived characters is the fact that these terms are relative. The same trait can be considered one or the other depending on the particular diagram being used. Returning to [\[link\]](#), note that the amniotic egg is a shared ancestral character for the Amniota clade, while having hair is a shared derived character for some organisms in this group. These terms help scientists distinguish between clades in the building of phylogenetic trees.

### Choosing the Right Relationships

Imagine being the person responsible for organizing all of the items in a department store properly—an overwhelming task. Organizing the evolutionary relationships of all life on Earth proves much more difficult: scientists must span enormous blocks of time and work with information from long-extinct organisms. Trying to decipher the proper connections, especially given the presence of homologies and analogies, makes the task of building an accurate tree of life extraordinarily difficult. Add to that the advancement of DNA technology, which now provides large quantities of genetic sequences to be used and analyzed. Taxonomy is a subjective

discipline: many organisms have more than one connection to each other, so each taxonomist will decide the order of connections.

To aid in the tremendous task of describing phylogenies accurately, scientists often use a concept called **maximum parsimony**, which means that events occurred in the simplest, most obvious way. For example, if a group of people entered a forest preserve to go hiking, based on the principle of maximum parsimony, one could predict that most of the people would hike on established trails rather than forge new ones.

For scientists deciphering evolutionary pathways, the same idea is used: the pathway of evolution probably includes the fewest major events that coincide with the evidence at hand. Starting with all of the homologous traits in a group of organisms, scientists look for the most obvious and simple order of evolutionary events that led to the occurrence of those traits.

**Note:**

Link to Learning



Head to this [website](#) to learn how maximum parsimony is used to create phylogenetic trees.

These tools and concepts are only a few of the strategies scientists use to tackle the task of revealing the evolutionary history of life on Earth.

Recently, newer technologies have uncovered surprising discoveries with unexpected relationships, such as the fact that people seem to be more closely related to fungi than fungi are to plants. Sound unbelievable? As the information about DNA sequences grows, scientists will become closer to mapping the evolutionary history of all life on Earth.



## Section Summary

To build phylogenetic trees, scientists must collect accurate information that allows them to make evolutionary connections between organisms. Using morphologic and molecular data, scientists work to identify homologous characteristics and genes. Similarities between organisms can stem either from shared evolutionary history (homologies) or from separate evolutionary paths (analogies). Newer technologies can be used to help distinguish homologies from analogies. After homologous information is identified, scientists use cladistics to organize these events as a means to determine an evolutionary timeline. Scientists apply the concept of maximum parsimony, which states that the order of events probably occurred in the most obvious and simple way with the least amount of steps. For evolutionary events, this would be the path with the least number of major divergences that correlate with the evidence.

## Art Connections

### Exercise:

#### Problem:

[\[link\]](#) Which animals in this figure belong to a clade that includes animals with hair? Which evolved first, hair or the amniotic egg?

---

#### Solution:

[\[link\]](#) Rabbits and humans belong in the clade that includes animals with hair. The amniotic egg evolved before hair because the Amniota clade is larger than the clade that encompasses animals with hair.

### Exercise:

**Problem:** [\[link\]](#) What is the largest clade in this diagram?

---

#### Solution:

[\[link\]](#) The largest clade encompasses the entire tree.

## Review Questions

### Exercise:

**Problem:** Which statement about analogies is correct?

- a. They occur only as errors.
- b. They are synonymous with homologous traits.
- c. They are derived by similar environmental constraints.
- d. They are a form of mutation.

---

**Solution:**

C

### Exercise:

**Problem:** What do scientists use to apply cladistics?

- a. homologous traits
- b. homoplasies
- c. analogous traits
- d. monophyletic groups

---

**Solution:**

A

### Exercise:

**Problem:**

What is true about organisms that are a part of the same clade?

- a. They all share the same basic characteristics.
- b. They evolved from a shared ancestor.

- c. They usually fall into the same classification taxa.
- d. They have identical phylogenies.

---

**Solution:**

B

**Exercise:**

**Problem:**

Why do scientists apply the concept of maximum parsimony?

- a. to decipher accurate phylogenies
- b. to eliminate analogous traits
- c. to identify mutations in DNA codes
- d. to locate homoplasies

---

**Solution:**

A

**Free Response**

**Exercise:**

**Problem:**

Dolphins and fish have similar body shapes. Is this feature more likely a homologous or analogous trait?

---

**Solution:**

Dolphins are mammals and fish are not, which means that their evolutionary paths (phylogenies) are quite separate. Dolphins probably adapted to have a similar body plan after returning to an aquatic lifestyle, and, therefore, this trait is probably analogous.

**Exercise:****Problem:**

Why is it so important for scientists to distinguish between homologous and analogous characteristics before building phylogenetic trees?

---

**Solution:**

Phylogenetic trees are based on evolutionary connections. If an analogous similarity were used on a tree, this would be erroneous and, furthermore, would cause the subsequent branches to be inaccurate.

**Exercise:**

**Problem:** Describe maximum parsimony.

---

**Solution:**

Maximum parsimony hypothesizes that events occurred in the simplest, most obvious way, and the pathway of evolution probably includes the fewest major events that coincide with the evidence at hand.

**Glossary**

analogy

(also, homoplasy) characteristic that is similar between organisms by convergent evolution, not due to the same evolutionary path

cladistics

system used to organize homologous traits to describe phylogenies

maximum parsimony

applying the simplest, most obvious way with the least number of steps

molecular systematics

technique using molecular evidence to identify phylogenetic relationships

monophyletic group

(also, clade) organisms that share a single ancestor

shared ancestral character

describes a characteristic on a phylogenetic tree that is shared by all organisms on the tree

shared derived character

describes a characteristic on a phylogenetic tree that is shared only by a certain clade of organisms

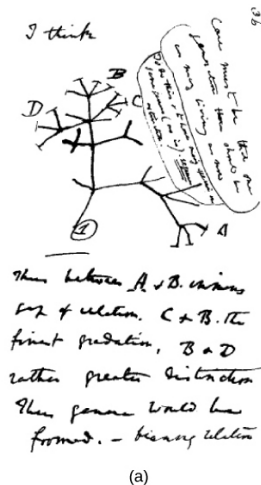
## Perspectives on the Phylogenetic Tree

By the end of this section, you will be able to:

- Describe horizontal gene transfer
- Illustrate how prokaryotes and eukaryotes transfer genes horizontally
- Identify the web and ring models of phylogenetic relationships and describe how they differ from the original phylogenetic tree concept

The concepts of phylogenetic modeling are constantly changing. It is one of the most dynamic fields of study in all of biology. Over the last several decades, new research has challenged scientists' ideas about how organisms are related. New models of these relationships have been proposed for consideration by the scientific community.

Many phylogenetic trees have been shown as models of the evolutionary relationship among species. Phylogenetic trees originated with Charles Darwin, who sketched the first phylogenetic tree in 1837 ([link](#)a), which served as a pattern for subsequent studies for more than a century. The concept of a phylogenetic tree with a single trunk representing a common ancestor, with the branches representing the divergence of species from this ancestor, fits well with the structure of many common trees, such as the oak ([link](#)b). However, evidence from modern DNA sequence analysis and newly developed computer algorithms has caused skepticism about the validity of the standard tree model in the scientific community.



The (a) concept of the “tree of life” goes back to an 1837 sketch by Charles Darwin. Like an (b) oak tree, the “tree of life” has a single trunk and many branches. (credit b:

modification of work by "Amada44"/Wikimedia Commons)

## **Limitations to the Classic Model**

Classical thinking about prokaryotic evolution, included in the classic tree model, is that species evolve clonally. That is, they produce offspring themselves with only random mutations causing the descent into the variety of modern-day and extinct species known to science. This view is somewhat complicated in eukaryotes that reproduce sexually, but the laws of Mendelian genetics explain the variation in offspring, again, to be a result of a mutation within the species. The concept of genes being transferred between unrelated species was not considered as a possibility until relatively recently. Horizontal gene transfer (HGT), also known as lateral gene transfer, is the transfer of genes between unrelated species. HGT has been shown to be an ever-present phenomenon, with many evolutionists postulating a major role for this process in evolution, thus complicating the simple tree model. Genes have been shown to be passed between species which are only distantly related using standard phylogeny, thus adding a layer of complexity to the understanding of phylogenetic relationships.

The various ways that HGT occurs in prokaryotes is important to understanding phylogenies. Although at present HGT is not viewed as important to eukaryotic evolution, HGT does occur in this domain as well. Finally, as an example of the ultimate gene transfer, theories of genome fusion between symbiotic or endosymbiotic organisms have been proposed to explain an event of great importance—the evolution of the first eukaryotic cell, without which humans could not have come into existence.

## **Horizontal Gene Transfer**

**Horizontal gene transfer (HGT)** is the introduction of genetic material from one species to another species by mechanisms other than the vertical transmission from parent(s) to offspring. These transfers allow even distantly related species to share genes, influencing their phenotypes. It is thought that HGT is more prevalent in prokaryotes, but that only about 2% of the

prokaryotic genome may be transferred by this process. Some researchers believe such estimates are premature: the actual importance of HGT to evolutionary processes must be viewed as a work in progress. As the phenomenon is investigated more thoroughly, it may be revealed to be more common. Many scientists believe that HGT and mutation appear to be (especially in prokaryotes) a significant source of genetic variation, which is the raw material for the process of natural selection. These transfers may occur between any two species that share an intimate relationship ([link](#)).

Summary of Mechanisms of Prokaryotic and Eukaryotic HGT			
	Mechanism	Mode of Transmission	Example
Prokaryotes	transformation	DNA uptake	many prokaryotes
	transduction	bacteriophage (virus)	bacteria
	conjugation	pilus	many prokaryotes
	gene transfer agents	phage-like particles	purple non-sulfur bacteria
Eukaryotes	from food organisms	unknown	aphid
	jumping genes	transposons	rice and millet plants



Summary of Mechanisms of Prokaryotic and Eukaryotic HGT			
	Mechanism	Mode of Transmission	Example
	epiphytes/parasites	unknown	yew tree fungi
	from viral infections		

## HGT in Prokaryotes

The mechanism of HGT has been shown to be quite common in the prokaryotic domains of Bacteria and Archaea, significantly changing the way their evolution is viewed. The majority of evolutionary models, such as in the Endosymbiont Theory, propose that eukaryotes descended from multiple prokaryotes, which makes HGT all the more important to understanding the phylogenetic relationships of all extant and extinct species.

The fact that genes are transferred among common bacteria is well known to microbiology students. These gene transfers between species are the major mechanism whereby bacteria acquire resistance to antibiotics. Classically, this type of transfer has been thought to occur by three different mechanisms:

1. Transformation: naked DNA is taken up by a bacteria
2. Transduction: genes are transferred using a virus
3. Conjugation: the use a hollow tube called a pilus to transfer genes between organisms

More recently, a fourth mechanism of gene transfer between prokaryotes has been discovered. Small, virus-like particles called **gene transfer agents (GTAs)** transfer random genomic segments from one species of prokaryote to another. GTAs have been shown to be responsible for genetic changes, sometimes at a very high frequency compared to other evolutionary processes. The first GTA was characterized in 1974 using purple, non-sulfur bacteria.

These GTAs, which are thought to be bacteriophages that lost the ability to reproduce on their own, carry random pieces of DNA from one organism to another. The ability of GTAs to act with high frequency has been demonstrated in controlled studies using marine bacteria. Gene transfer events in marine prokaryotes, either by GTAs or by viruses, have been estimated to be as high as  $10^{13}$  per year in the Mediterranean Sea alone. GTAs and viruses are thought to be efficient HGT vehicles with a major impact on prokaryotic evolution.

As a consequence of this modern DNA analysis, the idea that eukaryotes evolved directly from Archaea has fallen out of favor. While eukaryotes share many features that are absent in bacteria, such as the TATA box (found in the promoter region of many genes), the discovery that some eukaryotic genes were more homologous with bacterial DNA than Archaea DNA made this idea less tenable. Furthermore, the fusion of genomes from Archaea and Bacteria by endosymbiosis has been proposed as the ultimate event in eukaryotic evolution.

## **HGT in Eukaryotes**

Although it is easy to see how prokaryotes exchange genetic material by HGT, it was initially thought that this process was absent in eukaryotes. After all, prokaryotes are but single cells exposed directly to their environment, whereas the sex cells of multicellular organisms are usually sequestered in protected parts of the body. It follows from this idea that the gene transfers between multicellular eukaryotes should be more difficult. Indeed, it is thought that this process is rarer in eukaryotes and has a much smaller evolutionary impact than in prokaryotes. In spite of this fact, HGT between distantly related organisms has been demonstrated in several eukaryotic species, and it is possible that more examples will be discovered in the future.

In plants, gene transfer has been observed in species that cannot cross-pollinate by normal means. Transposons or “jumping genes” have been shown to transfer between rice and millet plant species. Furthermore, fungal species feeding on yew trees, from which the anti-cancer drug TAXOL® is derived from the bark, have acquired the ability to make taxol themselves, a clear example of gene transfer.

In animals, a particularly interesting example of HGT occurs within the aphid species ([\[link\]](#)). Aphids are insects that vary in color based on carotenoid content. Carotenoids are pigments made by a variety of plants, fungi, and microbes, and they serve a variety of functions in animals, who obtain these chemicals from their food. Humans require carotenoids to synthesize vitamin A, and we obtain them by eating orange fruits and vegetables: carrots, apricots, mangoes, and sweet potatoes. On the other hand, aphids have acquired the ability to make the carotenoids on their own. According to DNA analysis, this ability is due to the transfer of fungal genes into the insect by HGT, presumably as the insect consumed fungi for food. A carotenoid enzyme called a desaturase is responsible for the red coloration seen in certain aphids, and it has been further shown that when this gene is inactivated by mutation, the aphids revert back to their more common green color ([\[link\]](#)).



(a)



(b)

(a) Red aphids get their color from red carotenoid pigment. Genes necessary to make this pigment are present in certain fungi, and scientists speculate that aphids acquired these genes through HGT after consuming fungi for food. If genes for making carotenoids are inactivated by mutation, the aphids revert back to (b) their green color. Red coloration makes the aphids a lot more conspicuous to predators, but evidence suggests that red aphids are more resistant to insecticides than green ones. Thus, red aphids may be more fit to survive in some environments than green ones. (credit a: modification of work by Benny Mazur; credit b: modification of work by Mick Talbot)

## Genome Fusion and the Evolution of Eukaryotes

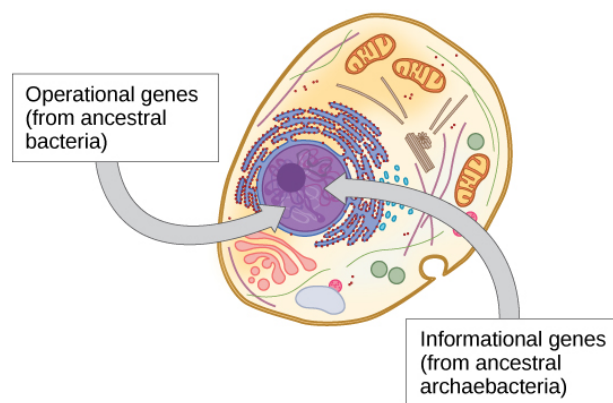
Scientists believe the ultimate in HGT occurs through **genome fusion** between different species of prokaryotes when two symbiotic organisms become endosymbiotic. This occurs when one species is taken inside the cytoplasm of another species, which ultimately results in a genome consisting of genes from both the endosymbiont and the host. This mechanism is an aspect of the Endosymbiont Theory, which is accepted by a majority of biologists as the mechanism whereby eukaryotic cells obtained their mitochondria and chloroplasts. However, the role of endosymbiosis in the development of the nucleus is more controversial. Nuclear and mitochondrial DNA are thought to be of different (separate) evolutionary origin, with the mitochondrial DNA being derived from the circular genomes of bacteria that were engulfed by ancient prokaryotic cells. Mitochondrial DNA can be regarded as the smallest chromosome. Interestingly enough, mitochondrial DNA is inherited only from the mother. The mitochondrial DNA degrades in sperm when the sperm degrades in the fertilized egg or in other instances when the mitochondria located in the flagellum of the sperm fails to enter the egg.

Within the past decade, the process of genome fusion by endosymbiosis has been proposed by James Lake of the UCLA/NASA Astrobiology Institute to be responsible for the evolution of the first eukaryotic cells ([\[link\]a](#)). Using DNA analysis and a new mathematical algorithm called conditioned reconstruction (CR), his laboratory proposed that eukaryotic cells developed from an endosymbiotic gene fusion between two species, one an Archaea and the other a Bacteria. As mentioned, some eukaryotic genes resemble those of Archaea, whereas others resemble those from Bacteria. An endosymbiotic fusion event, such as Lake has proposed, would clearly explain this observation. On the other hand, this work is new and the CR algorithm is relatively unsubstantiated, which causes many scientists to resist this hypothesis.

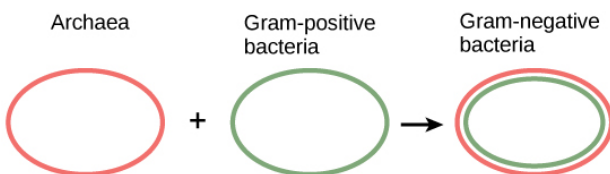
More recent work by Lake ([\[link\]b](#)) proposes that gram-negative bacteria, which are unique within their domain in that they contain two lipid bilayer membranes, indeed resulted from an endosymbiotic fusion of archaeal and bacterial species. The double membrane would be a direct result of the endosymbiosis, with the endosymbiont picking up the second membrane from

the host as it was internalized. This mechanism has also been used to explain the double membranes found in mitochondria and chloroplasts. Lake's work is not without skepticism, and the ideas are still debated within the biological science community. In addition to Lake's hypothesis, there are several other competing theories as to the origin of eukaryotes. How did the eukaryotic nucleus evolve? One theory is that the prokaryotic cells produced an additional membrane that surrounded the bacterial chromosome. Some bacteria have the DNA enclosed by two membranes; however, there is no evidence of a nucleolus or nuclear pores. Other proteobacteria also have membrane-bound chromosomes. If the eukaryotic nucleus evolved this way, we would expect one of the two types of prokaryotes to be more closely related to eukaryotes.

**(a) Genome fusion by endosymbiosis**



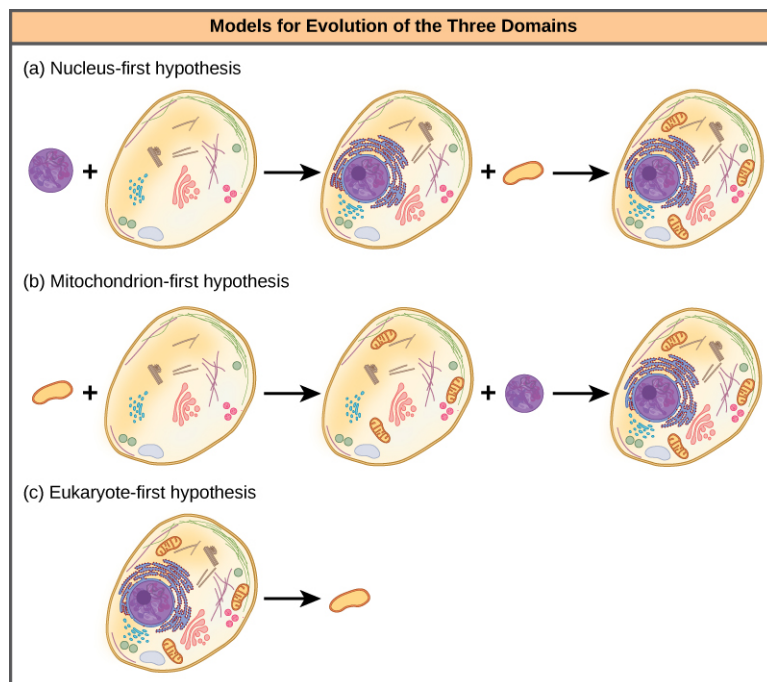
**(b) Endosymbiotic formation of Gram-negative bacteria**



The theory that mitochondria and chloroplasts are endosymbiotic in origin is now widely accepted. More controversial is the proposal that (a) the eukaryotic nucleus resulted from the fusion of archaeal and bacterial genomes, and that (b) Gram-negative

bacteria, which have two membranes, resulted from the fusion of Archaea and Gram-positive bacteria, each of which has a single membrane.

The **nucleus-first** hypothesis proposes that the nucleus evolved in prokaryotes first ([link]a), followed by a later fusion of the new eukaryote with bacteria that became mitochondria. The **mitochondria-first** hypothesis proposes that mitochondria were first established in a prokaryotic host ([link]b), which subsequently acquired a nucleus, by fusion or other mechanisms, to become the first eukaryotic cell. Most interestingly, the **eukaryote-first** hypothesis proposes that prokaryotes actually evolved from eukaryotes by losing genes and complexity ([link]c). All of these hypotheses are testable. Only time and more experimentation will determine which hypothesis is best supported by data.

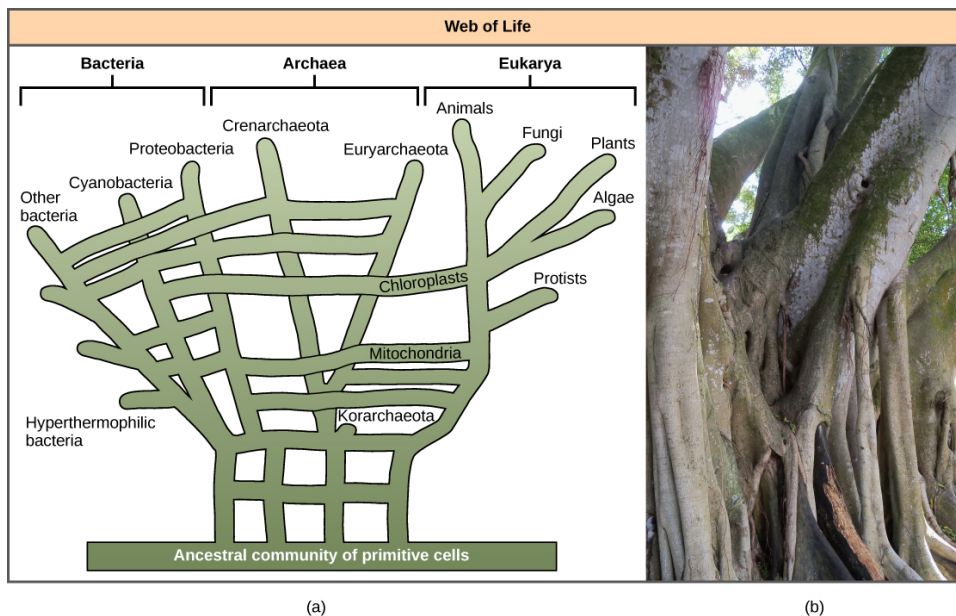


Three alternate hypotheses of eukaryotic and prokaryotic evolution are (a) the nucleus-

first hypothesis, (b) the mitochondrion-first hypothesis, and (c) the eukaryote-first hypothesis.

## Web and Network Models

The recognition of the importance of HGT, especially in the evolution of prokaryotes, has caused some to propose abandoning the classic “tree of life” model. In 1999, W. Ford Doolittle proposed a phylogenetic model that resembles a web or a network more than a tree. The hypothesis is that eukaryotes evolved not from a single prokaryotic ancestor, but from a pool of many species that were sharing genes by HGT mechanisms. As shown in [\[link\]a](#), some individual prokaryotes were responsible for transferring the bacteria that caused mitochondrial development to the new eukaryotes, whereas other species transferred the bacteria that gave rise to chloroplasts. This model is often called the “**web of life**.” In an effort to save the tree analogy, some have proposed using the *Ficus* tree ([\[link\]b](#)) with its multiple trunks as a phylogenetic to represent a diminished evolutionary role for HGT.



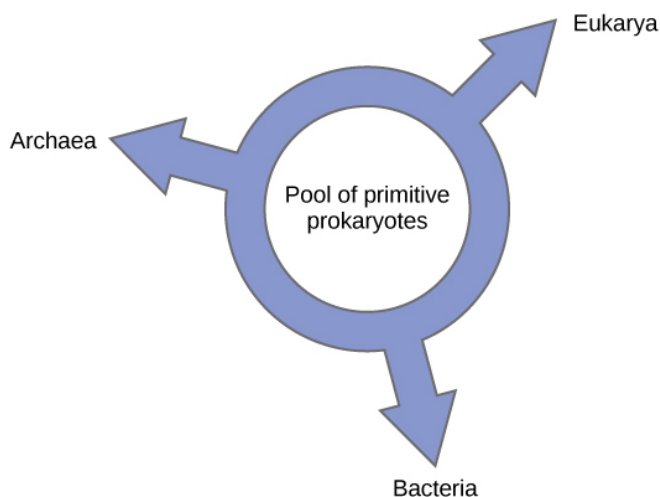
In the (a) phylogenetic model proposed by W. Ford



Doolittle, the “tree of life” arose from a community of ancestral cells, has multiple trunks, and has connections between branches where horizontal gene transfer has occurred. Visually, this concept is better represented by (b) the multi-trunked **Ficus** than by the single trunk of the oak similar to the tree drawn by Darwin [\[link\]](#).  
(credit b: modification of work by "psyberartist"/Flickr)

## Ring of Life Models

Others have proposed abandoning any tree-like model of phylogeny in favor of a ring structure, the so-called “**ring of life**” ([\[link\]](#)); a phylogenetic model where all three domains of life evolved from a pool of primitive prokaryotes. Lake, again using the conditioned reconstruction algorithm, proposes a ring-like model in which species of all three domains—Archaea, Bacteria, and Eukarya—evolved from a single pool of gene-swapping prokaryotes. His laboratory proposes that this structure is the best fit for data from extensive DNA analyses performed in his laboratory, and that the ring model is the only one that adequately takes HGT and genomic fusion into account. However, other phylogeneticists remain highly skeptical of this model.



According to the “ring of life”



phylogenetic model, the three domains of life evolved from a pool of primitive prokaryotes.

In summary, the “tree of life” model proposed by Darwin must be modified to include HGT. Does this mean abandoning the tree model completely? Even Lake argues that all attempts should be made to discover some modification of the tree model to allow it to accurately fit his data, and only the inability to do so will sway people toward his ring proposal.

This doesn’t mean a tree, web, or a ring will correlate completely to an accurate description of phylogenetic relationships of life. A consequence of the new thinking about phylogenetic models is the idea that Darwin’s original conception of the phylogenetic tree is too simple, but made sense based on what was known at the time. However, the search for a more useful model moves on: each model serving as hypotheses to be tested with the possibility of developing new models. This is how science advances. These models are used as visualizations to help construct hypothetical evolutionary relationships and understand the massive amount of data being analyzed.

## **Section Summary**

The phylogenetic tree, first used by Darwin, is the classic “tree of life” model describing phylogenetic relationships among species, and the most common model used today. New ideas about HGT and genome fusion have caused some to suggest revising the model to resemble webs or rings.

## **Review Questions**

### **Exercise:**

#### **Problem:**

The transfer of genes by a mechanism not involving asexual reproduction is called:

- a. meiosis

- b. web of life
- c. horizontal gene transfer
- d. gene fusion

---

**Solution:**

C

**Exercise:**

**Problem:**

Particles that transfer genetic material from one species to another, especially in marine prokaryotes:

- a. horizontal gene transfer
- b. lateral gene transfer
- c. genome fusion device
- d. gene transfer agents

---

**Solution:**

D

**Exercise:**

**Problem:** What does the trunk of the classic phylogenetic tree represent?

- a. single common ancestor
- b. pool of ancestral organisms
- c. new species
- d. old species

---

**Solution:**

A

**Exercise:**

**Problem:**

Which phylogenetic model proposes that all three domains of life evolved from a pool of primitive prokaryotes?

- a. tree of life
- b. web of life
- c. ring of life
- d. network model

---

**Solution:**

C

**Free Response****Exercise:****Problem:**

Compare three different ways that eukaryotic cells may have evolved.

---

**Solution:**

Some hypotheses propose that mitochondria were acquired first, followed by the development of the nucleus. Others propose that the nucleus evolved first and that this new eukaryotic cell later acquired the mitochondria. Still others hypothesize that prokaryotes descended from eukaryotes by the loss of genes and complexity.

**Exercise:**

**Problem:** Describe how aphids acquired the ability to change color.

---

**Solution:**

Aphids have acquired the ability to make the carotenoids on their own. DNA analysis has demonstrated that this ability is due to the transfer of

fungus genes into the insect by HGT, presumably as the insect consumed fungi for food.

## **Glossary**

eukaryote-first hypothesis

proposal that prokaryotes evolved from eukaryotes

gene transfer agent (GTA)

bacteriophage-like particle that transfers random genomic segments from one species of prokaryote to another

genome fusion

fusion of two prokaryotic genomes, presumably by endosymbiosis

horizontal gene transfer (HGT)

(also, lateral gene transfer) transfer of genes between unrelated species

mitochondria-first hypothesis

proposal that prokaryotes acquired a mitochondrion first, followed by nuclear development

nucleus-first hypothesis

proposal that prokaryotes acquired a nucleus first, and then the mitochondrion

ring of life

phylogenetic model where all three domains of life evolved from a pool of primitive prokaryotes

web of life

phylogenetic model that attempts to incorporate the effects of horizontal gene transfer on evolution

## Introduction

class="introduction"

Seedless  
plants, like  
these  
horsetails  
(*Equisetum*  
sp.), thrive in  
damp,  
shaded  
environment  
s under a tree  
canopy  
where  
dryness is  
rare. (credit:  
modification  
of work by  
Jerry  
Kirkhart)



An incredible variety of seedless plants populates the terrestrial landscape. Mosses may grow on a tree trunk, and horsetails may display their jointed stems and spindly leaves across the forest floor. Today, seedless plants represent only a small fraction of the plants in our environment; yet, three hundred million years ago, seedless plants dominated the landscape and grew in the enormous swampy forests of the Carboniferous period. Their decomposition created large deposits of coal that we mine today.

Current evolutionary thought holds that all plants—green algae as well as land dwellers—are monophyletic; that is, they are descendants of a single common ancestor. The evolutionary transition from water to land imposed severe constraints on plants. They had to develop strategies to avoid drying out, to disperse reproductive cells in air, for structural support, and for capturing and filtering sunlight. While seed plants developed adaptations that allowed them to populate even the most arid habitats on Earth, full independence from water did not happen in all plants. Most seedless plants still require a moist environment.

## Early Plant Life

By the end of this section, you will be able to:

- Discuss the challenges to plant life on land
- Describe the adaptations that allowed plants to colonize the land
- Describe the timeline of plant evolution and the impact of land plants on other living things

The kingdom Plantae constitutes large and varied groups of organisms. There are more than 300,000 species of catalogued plants. Of these, more than 260,000 are seed plants. Mosses, ferns, conifers, and flowering plants are all members of the plant kingdom. Most biologists also consider green algae to be plants, although others exclude all algae from the plant kingdom. The reason for this disagreement stems from the fact that only green algae, the **Charophytes**, share common characteristics with land plants (such as using chlorophyll *a* and *b* plus carotene in the same proportion as plants). These characteristics are absent in other types of algae.

### **Note:**

#### Evolution Connection

#### **Algae and Evolutionary Paths to Photosynthesis**

Some scientists consider all algae to be plants, while others assert that only the Charophytes belong in the kingdom Plantae. These divergent opinions are related to the different evolutionary paths to photosynthesis selected for in different types of algae. While all algae are photosynthetic—that is, they contain some form of a chloroplast—they didn't all become photosynthetic via the same path.

The ancestors to the green algae became photosynthetic by endosymbiosing a green, photosynthetic bacterium about 1.65 billion years ago. That algal line evolved into the Charophytes, and eventually into the modern mosses, ferns, gymnosperms, and angiosperms. Their evolutionary trajectory was relatively straight and monophyletic. In contrast, the other algae—red, brown, golden, stramenopiles, and so on—all became photosynthetic by secondary, or even tertiary, endosymbiotic events; that is, they endosymbiosed cells that had already endosymbiosed a

cyanobacterium. These latecomers to photosynthesis are parallels to the Charophytes in terms of autotrophy, but they did not expand to the same extent as the Charophytes, nor did they colonize the land.

The different views on whether all algae are Plantae arise from how these evolutionary paths are viewed. Scientists who solely track evolutionary straight lines (that is, monophyly), consider only the Charophytes as plants. To biologists who cast a broad net over living things that share a common characteristic (in this case, photosynthetic eukaryotes), all algae are plants.

**Note:**

Link to Learning



Go to this [interactive website](#) to get a more in-depth view of the Charophytes.

## Plant Adaptations to Life on Land

As organisms adapted to life on land, they had to contend with several challenges in the terrestrial environment. Water has been described as “the stuff of life.” The cell’s interior is a watery soup: in this medium, most small molecules dissolve and diffuse, and the majority of the chemical reactions of metabolism take place. Desiccation, or drying out, is a constant danger for an organism exposed to air. Even when parts of a plant are close to a source of water, the aerial structures are likely to dry out. Water also provides buoyancy to organisms. On land, plants need to develop structural support in a medium that does not give the same lift as water. The organism is also subject to bombardment by mutagenic radiation, because air does not



filter out ultraviolet rays of sunlight. Additionally, the male gametes must reach the female gametes using new strategies, because swimming is no longer possible. Therefore, both gametes and zygotes must be protected from desiccation. The successful land plants developed strategies to deal with all of these challenges. Not all adaptations appeared at once. Some species never moved very far from the aquatic environment, whereas others went on to conquer the driest environments on Earth.

To balance these survival challenges, life on land offers several advantages. First, sunlight is abundant. Water acts as a filter, altering the spectral quality of light absorbed by the photosynthetic pigment chlorophyll. Second, carbon dioxide is more readily available in air than in water, since it diffuses faster in air. Third, land plants evolved before land animals; therefore, until dry land was colonized by animals, no predators threatened plant life. This situation changed as animals emerged from the water and fed on the abundant sources of nutrients in the established flora. In turn, plants developed strategies to deter predation: from spines and thorns to toxic chemicals.

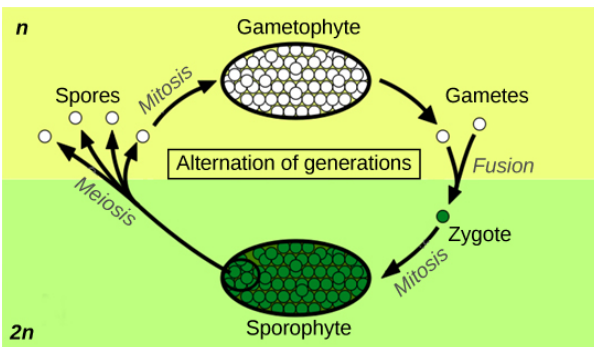
Early land plants, like the early land animals, did not live very far from an abundant source of water and developed survival strategies to combat dryness. One of these strategies is called tolerance. Many mosses, for example, can dry out to a brown and brittle mat, but as soon as rain or a flood makes water available, mosses will absorb it and are restored to their healthy green appearance. Another strategy is to colonize environments with high humidity, where droughts are uncommon. Ferns, which are considered an early lineage of plants, thrive in damp and cool places such as the understory of temperate forests. Later, plants moved away from moist or aquatic environments using resistance to desiccation, rather than tolerance. These plants, like cacti, minimize the loss of water to such an extent they can survive in extremely dry environments.

The most successful adaptation solution was the development of new structures that gave plants the advantage when colonizing new and dry environments. Four major adaptations are found in all terrestrial plants: the alternation of generations, a sporangium in which the spores are formed, a gametangium that produces haploid cells, and apical meristem tissue in

roots and shoots. The evolution of a waxy cuticle and a cell wall with lignin also contributed to the success of land plants. These adaptations are noticeably lacking in the closely related green algae—another reason for the debate over their placement in the plant kingdom.

## Alternation of Generations

Alternation of generations describes a life cycle in which an organism has both haploid and diploid multicellular stages ([link](#)).



Alternation of generations  
between the  $1n$  gametophyte  
and  $2n$  sporophyte is shown.  
(credit: Peter Coxhead)

**Haplontic** refers to a lifecycle in which there is a dominant haploid stage, and **diplontic** refers to a lifecycle in which the diploid is the dominant life stage. Humans are diplontic. Most plants exhibit alternation of generations, which is described as **haplodiplontic**: the haploid multicellular form, known as a gametophyte, is followed in the development sequence by a multicellular diploid organism: the sporophyte. The gametophyte gives rise to the gametes (reproductive cells) by mitosis. This can be the most obvious phase of the life cycle of the plant, as in the mosses, or it can occur in a

microscopic structure, such as a pollen grain, in the higher plants (a common collective term for the vascular plants). The sporophyte stage is barely noticeable in lower plants (the collective term for the plant groups of mosses, liverworts, and lichens). Towering trees are the diplontic phase in the lifecycles of plants such as sequoias and pines.

Protection of the embryo is a major requirement for land plants. The vulnerable embryo must be sheltered from desiccation and other environmental hazards. In both seedless and seed plants, the female gametophyte provides protection and nutrients to the embryo as it develops into the new generation of sporophyte. This distinguishing feature of land plants gave the group its alternate name of **embryophytes**.

### **Sporangia in Seedless Plants**

The sporophyte of seedless plants is diploid and results from syngamy (fusion) of two gametes. The sporophyte bears the sporangia (singular, sporangium): organs that first appeared in the land plants. The term “sporangia” literally means “spore in a vessel,” as it is a reproductive sac that contains spores [\[link\]](#). Inside the multicellular sporangia, the diploid **sporocytes**, or mother cells, produce haploid spores by meiosis, where the  $2n$  chromosome number is reduced to  $1n$  (note that many plant sporophytes are polyploid: for example, durum wheat is tetraploid, bread wheat is hexaploid, and some ferns are 1000-ploid). The spores are later released by the sporangia and disperse in the environment. Two different types of spores are produced in land plants, resulting in the separation of sexes at different points in the lifecycle. **Seedless non-vascular plants** produce only one kind of spore and are called **homosporous**. The gametophyte phase is dominant in these plants. After germinating from a spore, the resulting gametophyte produces both male and female gametangia, usually on the same individual. In contrast, **heterosporous** plants produce two morphologically different types of spores. The male spores are called **microspores**, because of their smaller size, and develop into the male gametophyte; the comparatively larger **megaspores** develop into the female gametophyte. Heterospory is observed in a few **seedless vascular plants** and in all seed plants.



Spore-producing sacs called sporangia grow at the ends of long, thin stalks in this photo of the moss *Esporangios bryum*.  
(credit: Javier Martin)

When the haploid spore germinates in a hospitable environment, it generates a multicellular gametophyte by mitosis. The gametophyte supports the zygote formed from the fusion of gametes and the resulting young sporophyte (vegetative form). The cycle then begins anew.

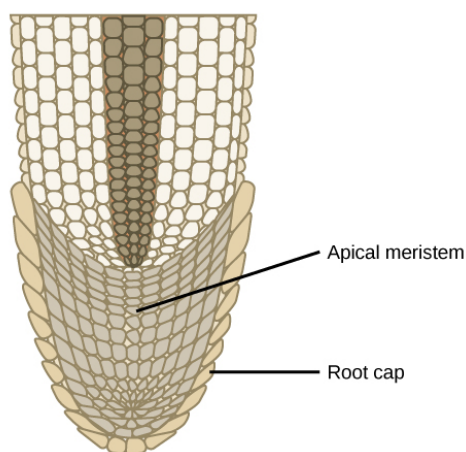
The spores of seedless plants are surrounded by thick cell walls containing a tough polymer known as **sporopollenin**. This complex substance is characterized by long chains of organic molecules related to fatty acids and carotenoids: hence the yellow color of most pollen. Sporopollenin is unusually resistant to chemical and biological degradation. In seed plants, which use pollen to transfer the male sperm to the female egg, the toughness of sporopollenin explains the existence of well-preserved pollen fossils. Sporopollenin was once thought to be an innovation of land plants; however, the green algae *Coleochaetes* forms spores that contain sporopollenin.

## Gametangia in Seedless Plants

**Gametangia** (singular, gametangium) are structures observed on multicellular haploid gametophytes. In the gametangia, precursor cells give rise to gametes by mitosis. The male gametangium (**antheridium**) releases sperm. Many seedless plants produce sperm equipped with flagella that enable them to swim in a moist environment to the **archegonia**: the female gametangium. The embryo develops inside the archegonium as the sporophyte. Gametangia are prominent in seedless plants, but are very rarely found in seed plants.

## Apical Meristems

Shoots and roots of plants increase in length through rapid cell division in a tissue called the apical meristem, which is a small zone of cells found at the shoot tip or root tip ([link](#)). The apical meristem is made of undifferentiated cells that continue to proliferate throughout the life of the plant. Meristematic cells give rise to all the specialized tissues of the organism. Elongation of the shoots and roots allows a plant to access additional space and resources: light in the case of the shoot, and water and minerals in the case of roots. A separate meristem, called the lateral meristem, produces cells that increase the diameter of tree trunks.



Addition of new cells in a root occurs at the apical

meristem. Subsequent enlargement of these cells causes the organ to grow and elongate. The root cap protects the fragile apical meristem as the root tip is pushed through the soil by cell elongation.

## **Additional Land Plant Adaptations**

As plants adapted to dry land and became independent from the constant presence of water in damp habitats, new organs and structures made their appearance. Early land plants did not grow more than a few inches off the ground, competing for light on these low mats. By developing a shoot and growing taller, individual plants captured more light. Because air offers substantially less support than water, land plants incorporated more rigid molecules in their stems (and later, tree trunks). In small plants such as single-celled algae, simple diffusion suffices to distribute water and nutrients throughout the organism. However, for plants to evolve larger forms, the evolution of vascular tissue for the distribution of water and solutes was a prerequisite. The vascular system contains xylem and phloem tissues. Xylem conducts water and minerals absorbed from the soil up to the shoot, while phloem transports food derived from photosynthesis throughout the entire plant. A root system evolved to take up water and minerals from the soil, and to anchor the increasingly taller shoot in the soil.

In land plants, a waxy, waterproof cover called a cuticle protects the leaves and stems from desiccation. However, the cuticle also prevents intake of carbon dioxide needed for the synthesis of carbohydrates through photosynthesis. To overcome this, stomata or pores that open and close to regulate traffic of gases and water vapor appeared in plants as they moved away from moist environments into drier habitats.

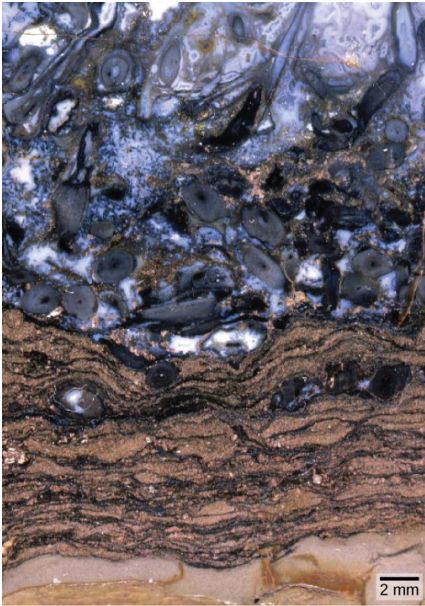
Water filters ultraviolet-B (UVB) light, which is harmful to all organisms, especially those that must absorb light to survive. This filtering does not occur for land plants. This presented an additional challenge to land colonization, which was met by the evolution of biosynthetic pathways for the synthesis of protective flavonoids and other compounds: pigments that absorb UV wavelengths of light and protect the aerial parts of plants from photodynamic damage.

Plants cannot avoid being eaten by animals. Instead, they synthesize a large range of poisonous secondary metabolites: complex organic molecules such as alkaloids, whose noxious smells and unpleasant taste deter animals. These toxic compounds can also cause severe diseases and even death, thus discouraging predation. Humans have used many of these compounds for centuries as drugs, medications, or spices. In contrast, as plants co-evolved with animals, the development of sweet and nutritious metabolites lured animals into providing valuable assistance in dispersing pollen grains, fruit, or seeds. Plants have been enlisting animals to be their helpers in this way for hundreds of millions of years.

## **Evolution of Land Plants**

No discussion of the evolution of plants on land can be undertaken without a brief review of the timeline of the geological eras. The early era, known as the Paleozoic, is divided into six periods. It starts with the Cambrian period, followed by the Ordovician, Silurian, Devonian, Carboniferous, and Permian. The major event to mark the Ordovician, more than 500 million years ago, was the colonization of land by the ancestors of modern land plants. Fossilized cells, cuticles, and spores of early land plants have been dated as far back as the Ordovician period in the early Paleozoic era. The oldest-known vascular plants have been identified in deposits from the Devonian. One of the richest sources of information is the Rhynie chert, a sedimentary rock deposit found in Rhynie, Scotland ([link](#)), where embedded fossils of some of the earliest vascular plants have been identified.





(a)



(b)

This Rhynie chert contains fossilized material from vascular plants. The area inside the circle contains bulbous underground stems called corms, and root-like structures called rhizoids. (credit b: modification of work by Peter Coxhead based on original image by “Smith609”/Wikimedia Commons; scale-bar data from Matt Russell)

Paleobotanists distinguish between **extinct** species, as fossils, and **extant** species, which are still living. The extinct vascular plants, classified as zosterophylls and trimerophytes, most probably lacked true leaves and roots and formed low vegetation mats similar in size to modern-day mosses, although some trimetophytes could reach one meter in height. The later genus *Cooksonia*, which flourished during the Silurian, has been extensively studied from well-preserved examples. Imprints of *Cooksonia* show slender branching stems ending in what appear to be sporangia. From the recovered specimens, it is not possible to establish for certain whether *Cooksonia* possessed vascular tissues. Fossils indicate that by the end of the Devonian period, ferns, horsetails, and seed plants populated the landscape, giving rising to trees and forests. This luxuriant vegetation helped enrich



the atmosphere in oxygen, making it easier for air-breathing animals to colonize dry land. Plants also established early symbiotic relationships with fungi, creating mycorrhizae: a relationship in which the fungal network of filaments increases the efficiency of the plant root system, and the plants provide the fungi with byproducts of photosynthesis.

**Note:**

**Career Connection**

**Paleobotanist**

How organisms acquired traits that allow them to colonize new environments—and how the contemporary ecosystem is shaped—are fundamental questions of evolution. Paleobotany (the study of extinct plants) addresses these questions through the analysis of fossilized specimens retrieved from field studies, reconstituting the morphology of organisms that disappeared long ago. Paleobotanists trace the evolution of plants by following the modifications in plant morphology: shedding light on the connection between existing plants by identifying common ancestors that display the same traits. This field seeks to find transitional species that bridge gaps in the path to the development of modern organisms. Fossils are formed when organisms are trapped in sediments or environments where their shapes are preserved. Paleobotanists collect fossil specimens in the field and place them in the context of the geological sediments and other fossilized organisms surrounding them. The activity requires great care to preserve the integrity of the delicate fossils and the layers of rock in which they are found.

One of the most exciting recent developments in paleobotany is the use of analytical chemistry and molecular biology to study fossils. Preservation of molecular structures requires an environment free of oxygen, since oxidation and degradation of material through the activity of microorganisms depend on its presence. One example of the use of analytical chemistry and molecular biology is the identification of oleanane, a compound that deters pests. Up to this point, oleanane appeared to be unique to flowering plants; however, it has now been recovered from sediments dating from the Permian, much earlier than the current dates given for the appearance of the first flowering plants.

Paleobotanists can also study fossil DNA, which can yield a large amount of information, by analyzing and comparing the DNA sequences of extinct plants with those of living and related organisms. Through this analysis, evolutionary relationships can be built for plant lineages.

Some paleobotanists are skeptical of the conclusions drawn from the analysis of molecular fossils. For example, the chemical materials of interest degrade rapidly when exposed to air during their initial isolation, as well as in further manipulations. There is always a high risk of contaminating the specimens with extraneous material, mostly from microorganisms. Nevertheless, as technology is refined, the analysis of DNA from fossilized plants will provide invaluable information on the evolution of plants and their adaptation to an ever-changing environment.

## The Major Divisions of Land Plants

The green algae and land plants are grouped together into a subphylum called the Streptophytina, and thus are called Streptophytes. In a further division, land plants are classified into two major groups according to the absence or presence of vascular tissue, as detailed in [\[link\]](#). Plants that lack vascular tissue, which is formed of specialized cells for the transport of water and nutrients, are referred to as **non-vascular plants**. Liverworts, mosses, and hornworts are seedless, non-vascular plants that likely appeared early in land plant evolution. Vascular plants developed a network of cells that conduct water and solutes. The first vascular plants appeared in the late Ordovician and were probably similar to lycophytes, which include club mosses (not to be confused with the mosses) and the pterophytes (ferns, horsetails, and whisk ferns). Lycophytes and pterophytes are referred to as seedless vascular plants, because they do not produce seeds. The seed plants, or spermatophytes, form the largest group of all existing plants, and hence dominate the landscape. Seed plants include gymnosperms, most notably conifers (Gymnosperms), which produce “naked seeds,” and the most successful of all plants, the flowering plants (Angiosperms). Angiosperms protect their seeds inside chambers at the center of a flower; the walls of the chamber later develop into a fruit.

**Note:****Art Connection**

STREPTOPHYTES: THE GREEN PLANTS							
Charophytes	Embryophytes: The Land Plants						
	Non Vascular			Vascular			
	Seedless Plants Bryophytes			Seedless Plants		Seed Plants Spermatophytes	
				Lycophytes	Pterophytes		
	Liver- worts	Horn- worts	Mosses	Club Mosses	Whisk Ferns	Gymno- sperms	Angio- sperms
				Quillworts	Horsetails		

This table shows the major divisions of green plants.

Which of the following statements about plant divisions is false?

- a. Lycophytes and pterophytes are seedless vascular plants.
- b. All vascular plants produce seeds.
- c. All nonvascular embryophytes are bryophytes.
- d. Seed plants include angiosperms and gymnosperms.

**Section Summary**

Land plants acquired traits that made it possible to colonize land and survive out of the water. All land plants share the following characteristics: alternation of generations, with the haploid plant called a gametophyte, and the diploid plant called a sporophyte; protection of the embryo, formation of haploid spores in a sporangium, formation of gametes in a gametangium, and an apical meristem. Vascular tissues, roots, leaves, cuticle cover, and a tough outer layer that protects the spores contributed to the adaptation of plants to dry land. Land plants appeared about 500 million years ago in the Ordovician period.

## Art Connections

### Exercise:

#### Problem:

[\[link\]](#) Which of the following statements about plant divisions is false?

- a. Lycophytes and pterophytes are seedless vascular plants.
- b. All vascular plants produce seeds.
- c. All nonvascular embryophytes are bryophytes.
- d. Seed plants include angiosperms and gymnosperms.

---

#### Solution:

[\[link\]](#) B.

## Review Questions

### Exercise:

#### Problem:

The land plants are probably descendants of which of these groups?

- a. green algae
- b. red algae
- c. brown algae
- d. angiosperms

---

#### Solution:

A

### Exercise:

**Problem:** Alternation of generations means that plants produce:

- a. only haploid multicellular organisms
  - b. only diploid multicellular organisms
  - c. only diploid multicellular organisms with single-celled haploid gametes
  - d. both haploid and diploid multicellular organisms
- 

**Solution:**

D

**Exercise:**

**Problem:**

Which of the following traits of land plants allows them to grow in height?

- a. alternation of generations
  - b. waxy cuticle
  - c. tracheids
  - d. sporopollenin
- 

**Solution:**

C

**Free Response**

**Exercise:**

**Problem:**

Why did land plants lose some of the accessory pigments present in brown and red algae?

---

**Solution:**

Sunlight is not filtered by water or other algae on land; therefore, there is no need to collect light at additional wavelengths made available by other pigment coloration.

**Exercise:**

**Problem:** What is the difference between extant and extinct?

---

**Solution:**

Paleobotanists distinguish between extinct species, which no longer live, and extant species, which are still living.

**Glossary**

antheridium

male gametangium

archegonium

female gametangium

charophyte

other term for green algae; considered the closest relative of land plants

diplontic

diploid stage is the dominant stage

embryophyte

other name for land plant; embryo is protected and nourished by the sporophyte

extant

still-living species

extinct

no longer existing species

gametangium

structure on the gametophyte in which gametes are produced

haplodiplodontic

haploid and diploid stages alternate

haplontic

haploid stage is the dominant stage

heterosporous

produces two types of spores

homosporous

produces one type of spore

megaspore

female spore

microspore

male spore

non-vascular plant

plant that lacks vascular tissue, which is formed of specialized cells for the transport of water and nutrients

seedless vascular plant

plant that does not produce seeds

sporocyte

diploid cell that produces spores by meiosis

sporopollenin

tough polymer surrounding the spore

vascular plant

plant containing a network of cells that conducts water and solutes through the organism

## Green Algae: Precursors of Land Plants

By the end of this section, you will be able to:

- Describe the traits shared by green algae and land plants
- Explain the reasons why Charales are considered the closest relative to land plants
- Understand that current phylogenetic relationships are reshaped by comparative analysis of DNA sequences

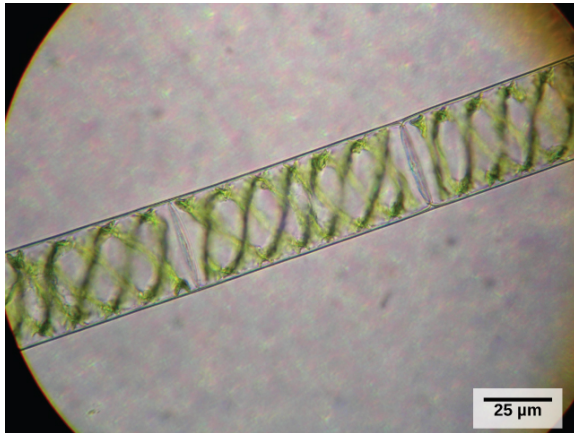
## Streptophytes

Until recently, all photosynthetic eukaryotes were considered members of the kingdom Plantae. The brown, red, and gold algae, however, have been reassigned to the Protista kingdom. This is because apart from their ability to capture light energy and fix CO<sub>2</sub>, they lack many structural and biochemical traits that distinguish plants from protists. The position of green algae is more ambiguous. Green algae contain the same carotenoids and chlorophyll *a* and *b* as land plants, whereas other algae have different accessory pigments and types of chlorophyll molecules in addition to chlorophyll *a*. Both green algae and land plants also store carbohydrates as starch. Cells in green algae divide along cell plates called phragmoplasts, and their cell walls are layered in the same manner as the cell walls of embryophytes. Consequently, land plants and closely related green algae are now part of a new monophyletic group called **Streptophyta**.

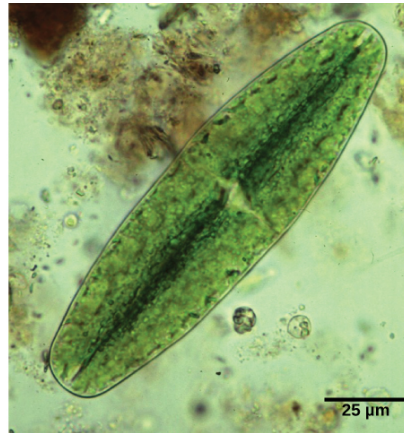
The remaining green algae, which belong to a group called Chlorophyta, include more than 7000 different species that live in fresh or brackish water, in seawater, or in snow patches. A few green algae even survive on soil, provided it is covered by a thin film of moisture in which they can live. Periodic dry spells provide a selective advantage to algae that can survive water stress. Some green algae may already be familiar, in particular *Spirogyra* and desmids. Their cells contain chloroplasts that display a dizzying variety of shapes, and their cell walls contain cellulose, as do land plants. Some green algae are single cells, such as *Chlorella* and *Chlamydomonas*, which adds to the ambiguity of green algae classification, because plants are multicellular. Other algae, like *Ulva* (commonly called sea lettuce), form colonies ([\[link\]](#)).



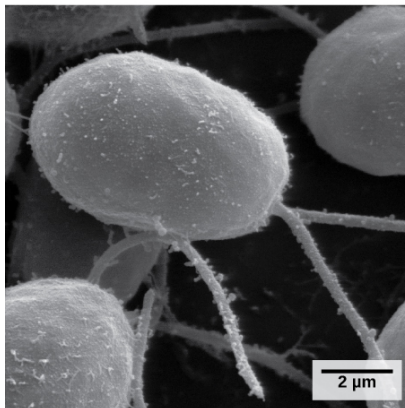
Chlorophyta include (a) *Spirogyra*, (b) desmids, (c) *Chlamydomonas*, and (d) *Ulva*. Desmids and *Chlamydomonas* are single-celled organisms, *Spirogyra* forms chains of cells, and *Ulva* forms colonies resembling leaves (credit b: modification of work by Derek Keats; credit c: modification of work by Dartmouth Electron Microscope Facility, Dartmouth College; credit d: modification of work by Holger Krisp; scale-bar data from Matt Russell)



(a) *Spirogyra*



(b) Desmid



(c) *Chlamydomonas*



(d) *Ulva*

## Reproduction of Green Algae

Green algae reproduce both asexually, by fragmentation or dispersal of spores, or sexually, by producing gametes that fuse during fertilization. In a single-celled organism such as *Chlamydomonas*, there is no mitosis after

fertilization. In the multicellular *Ulva*, a sporophyte grows by mitosis after fertilization. Both *Chlamydomonas* and *Ulva* produce flagellated gametes.

## Charales

Green algae in the order Charales, and the coleochaetes (microscopic green algae that enclose their spores in sporopollenin), are considered the closest living relatives of embryophytes. The Charales can be traced back 420 million years. They live in a range of fresh water habitats and vary in size from a few millimeters to a meter in length. The representative species is *Chara* ([\[link\]](#)), often called muskgrass or skunkweed because of its unpleasant smell. Large cells form the thallus: the main stem of the alga. Branches arising from the nodes are made of smaller cells. Male and female reproductive structures are found on the nodes, and the sperm have flagella. Unlike land plants, Charales do not undergo alternation of generations in their lifecycle. Charales exhibit a number of traits that are significant in their adaptation to land life. They produce the compounds lignin and sporopollenin, and form plasmodesmata that connect the cytoplasm of adjacent cells. The egg, and later, the zygote, form in a protected chamber on the parent plant.



The representative alga, *Chara*, is a noxious weed in Florida, where it clogs waterways. (credit: South

New information from recent, extensive DNA sequence analysis of green algae indicates that the Zygnematales are more closely related to the embryophytes than the Charales. The Zygnematales include the familiar genus *Spirogyra*. As techniques in DNA analysis improve and new information on comparative genomics arises, the phylogenetic connections between species will change. Clearly, plant biologists have not yet solved the mystery of the origin of land plants.

## Section Summary

Green algae share more traits with land plants than other algae, according to structure and DNA analysis. Charales form sporopollenin and precursors of lignin, phragmoplasts, and have flagellated sperm. They do not exhibit alternation of generations.

## Review Questions

### Exercise:

#### Problem:

What characteristic of Charales would enable them to survive a dry spell?

- a. sperm with flagella
- b. phragmoplasts
- c. sporopollenin
- d. chlorophyll *a*

---

### Solution:

C

**Exercise:**

**Problem:**

Which one of these characteristics is present in land plants and not in Charales?

- a. alternation of generations
- b. flagellated sperm
- c. phragmoplasts
- d. plasmodesmata

---

**Solution:**

A

**Free Response**

**Exercise:**

**Problem:**

To an alga, what is the main advantage of producing drought-resistant structures?

---

**Solution:**

It allows for survival through periodic droughts and colonization of environments where the supply of water fluctuates.

**Glossary**

streptophytes

group that includes green algae and land plants

## Bryophytes

By the end of this section, you will be able to:

- Identify the main characteristics of bryophytes
- Describe the distinguishing traits of liverworts, hornworts, and mosses
- Chart the development of land adaptations in the bryophytes
- Describe the events in the bryophyte lifecycle

Bryophytes are the group of plants that are the closest extant relative of early terrestrial plants. The first bryophytes (liverworts) most likely appeared in the Ordovician period, about 450 million years ago. Because of the lack of lignin and other resistant structures, the likelihood of bryophytes forming fossils is rather small. Some spores protected by sporopollenin have survived and are attributed to early bryophytes. By the Silurian period, however, vascular plants had spread through the continents. This compelling fact is used as evidence that non-vascular plants must have preceded the Silurian period.

More than 25,000 species of bryophytes thrive in mostly damp habitats, although some live in deserts. They constitute the major flora of inhospitable environments like the tundra, where their small size and tolerance to desiccation offer distinct advantages. They generally lack lignin and do not have actual tracheids (xylem cells specialized for water conduction). Rather, water and nutrients circulate inside specialized conducting cells. Although the term non-tracheophyte is more accurate, bryophytes are commonly called nonvascular plants.

In a bryophyte, all the conspicuous vegetative organs—including the photosynthetic leaf-like structures, the thallus, stem, and the rhizoid that anchors the plant to its substrate—belong to the haploid organism or gametophyte. The sporophyte is barely noticeable. The gametes formed by bryophytes swim with a flagellum, as do gametes in a few of the tracheophytes. The sporangium—the multicellular sexual reproductive structure—is present in bryophytes and absent in the majority of algae. The bryophyte embryo also remains attached to the parent plant, which protects and nourishes it. This is a characteristic of land plants.

The bryophytes are divided into three phyla: the liverworts or Hepaticophyta, the hornworts or Anthocerotophyta, and the mosses or true Bryophyta.

## Liverworts

**Liverworts** (Hepaticophyta) are viewed as the plants most closely related to the ancestor that moved to land. Liverworts have colonized every terrestrial habitat on Earth and diversified to more than 7000 existing species ([link](#)). Some gametophytes form lobate green structures, as seen in [link](#). The shape is similar to the lobes of the liver, and hence provides the origin of the name given to the phylum.



This 1904 drawing shows the variety of forms of Hepaticophyta.

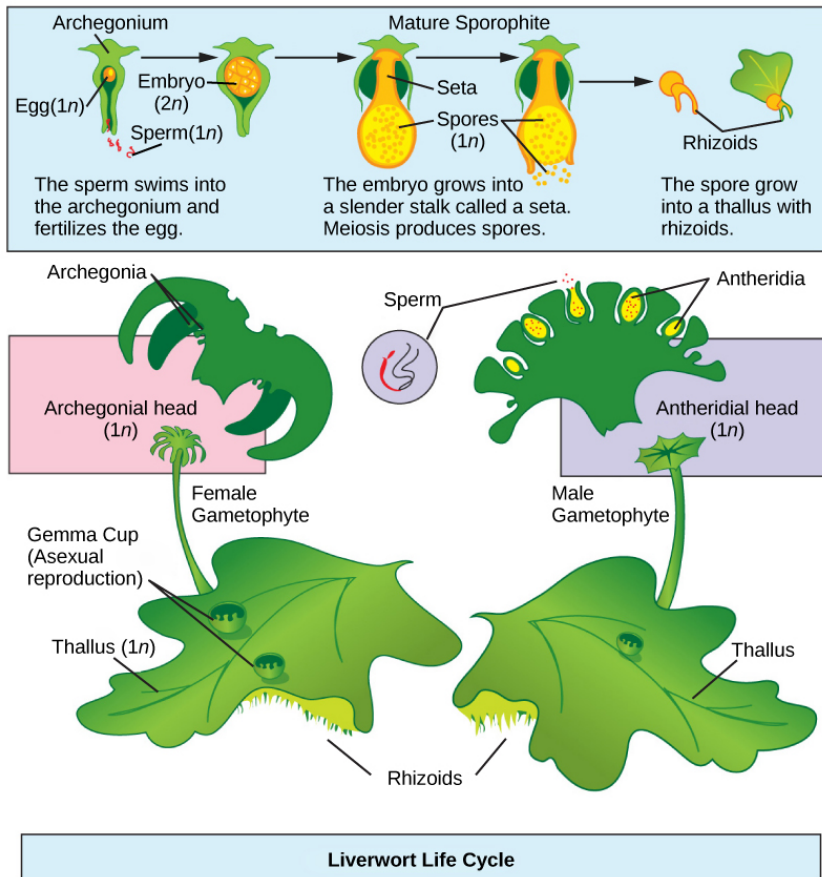




A liverwort, *Lunularia cruciata*, displays its lobate, flat thallus. The organism in the photograph is in the gametophyte stage.

Openings that allow the movement of gases may be observed in liverworts. However, these are not stomata, because they do not actively open and close. The plant takes up water over its entire surface and has no cuticle to prevent desiccation. [\[link\]](#) represents the lifecycle of a liverwort. The cycle starts with the release of haploid spores from the sporangium that developed on the sporophyte. Spores disseminated by wind or water germinate into flattened thalli attached to the substrate by thin, single-celled filaments. Male and female gametangia develop on separate, individual plants. Once released, male gametes swim with the aid of their flagella to the female gametangium (the archegonium), and fertilization ensues. The zygote grows into a small sporophyte still attached to the parent gametophyte. It will give rise, by meiosis, to the next generation of spores. Liverwort plants can also reproduce asexually, by the breaking of branches or the spreading of leaf fragments called gemmae. In this latter type of reproduction, the **gemmae**—small, intact, complete pieces of plant that are produced in a cup on the

surface of the thallus (shown in [link](#))—are splashed out of the cup by raindrops. The gemmae then land nearby and develop into gametophytes.



The life cycle of a typical liverwort is shown.  
(credit: modification of work by Mariana Ruiz Villareal)

## Hornworts

The **hornworts** (*Anthocerotophyta*) belong to the broad bryophyte group. They have colonized a variety of habitats on land, although they are never far from a source of moisture. The short, blue-green gametophyte is the dominant phase of the lifecycle of a hornwort. The narrow, pipe-like



sporophyte is the defining characteristic of the group. The sporophytes emerge from the parent gametophyte and continue to grow throughout the life of the plant ([link](#)).

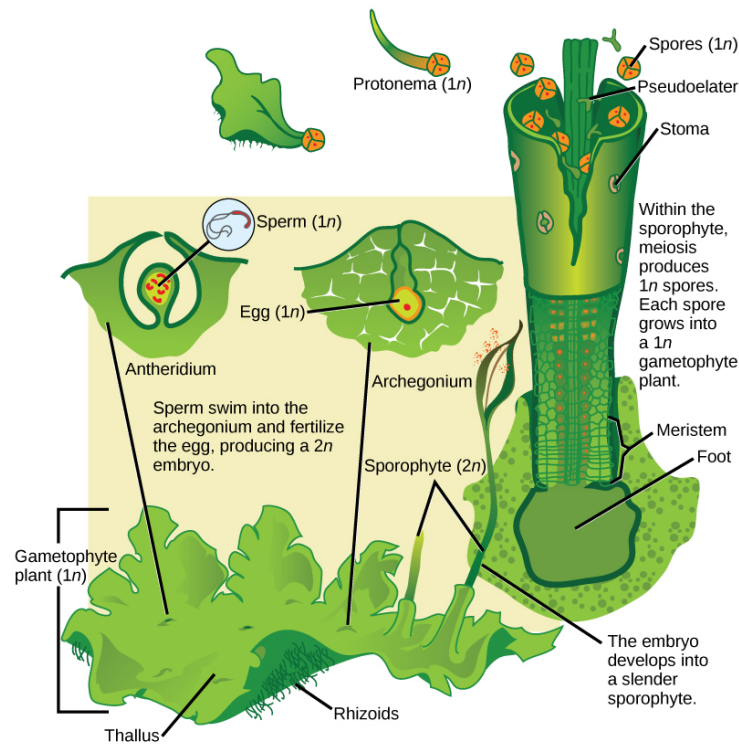


Hornworts grow a tall and slender sporophyte. (credit: modification of work by Jason Hollinger)

Stomata appear in the hornworts and are abundant on the sporophyte. Photosynthetic cells in the thallus contain a single chloroplast. Meristem cells at the base of the plant keep dividing and adding to its height. Many hornworts establish symbiotic relationships with cyanobacteria that fix nitrogen from the environment.

The lifecycle of hornworts ([link](#)) follows the general pattern of alternation of generations. The gametophytes grow as flat thalli on the soil with embedded gametangia. Flagellated sperm swim to the archegonia and fertilize eggs. The zygote develops into a long and slender sporophyte that eventually splits open, releasing spores. Thin cells called pseudoelaters surround the spores and help propel them further in the environment. Unlike the elaters observed in horsetails, the hornwort pseudoelaters are

single-celled structures. The haploid spores germinate and give rise to the next generation of gametophyte.



The alternation of generation in hornworts is shown. (credit: modification of work by "Smith609"/Wikimedia Commons based on original work by Mariana Ruiz Villareal)

## Mosses

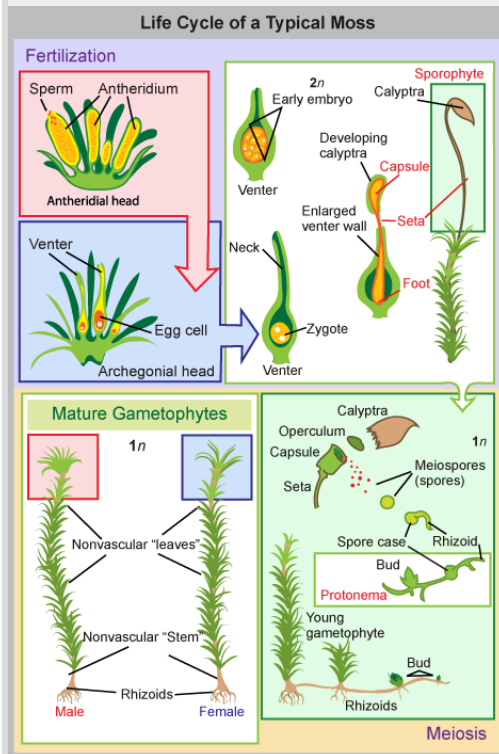
More than 10,000 species of **mosses** have been catalogued. Their habitats vary from the tundra, where they are the main vegetation, to the understory of tropical forests. In the tundra, the mosses' shallow rhizoids allow them to fasten to a substrate without penetrating the frozen soil. Mosses slow down

erosion, store moisture and soil nutrients, and provide shelter for small animals as well as food for larger herbivores, such as the musk ox. Mosses are very sensitive to air pollution and are used to monitor air quality. They are also sensitive to copper salts, so these salts are a common ingredient of compounds marketed to eliminate mosses from lawns.

Mosses form diminutive gametophytes, which are the dominant phase of the lifecycle. Green, flat structures—resembling true leaves, but lacking vascular tissue—are attached in a spiral to a central stalk. The plants absorb water and nutrients directly through these leaf-like structures. Some mosses have small branches. Some primitive traits of green algae, such as flagellated sperm, are still present in mosses that are dependent on water for reproduction. Other features of mosses are clearly adaptations to dry land. For example, stomata are present on the stems of the sporophyte, and a primitive vascular system runs up the sporophyte's stalk. Additionally, mosses are anchored to the substrate—whether it is soil, rock, or roof tiles—by multicellular **rhizoids**. These structures are precursors of roots. They originate from the base of the gametophyte, but are not the major route for the absorption of water and minerals. The lack of a true root system explains why it is so easy to rip moss mats from a tree trunk. The moss lifecycle follows the pattern of alternation of generations as shown in [\[link\]](#). The most familiar structure is the haploid gametophyte, which germinates from a haploid spore and forms first a **protonema**—usually, a tangle of single-celled filaments that hug the ground. Cells akin to an apical meristem actively divide and give rise to a gametophore, consisting of a photosynthetic stem and foliage-like structures. Rhizoids form at the base of the gametophore. Gametangia of both sexes develop on separate gametophores. The male organ (the antheridium) produces many sperm, whereas the archegonium (the female organ) forms a single egg. At fertilization, the sperm swims down the neck to the venter and unites with the egg inside the archegonium. The zygote, protected by the archegonium, divides and grows into a sporophyte, still attached by its foot to the gametophyte.

**Note:**

## Art Connection



This illustration shows the life cycle of mosses. (credit: modification of work by Mariana Ruiz Villareal)

Which of the following statements about the moss life cycle is false?

- The mature gametophyte is haploid.
- The sporophyte produces haploid spores.
- The calyptra buds to form a mature gametophyte.
- The zygote is housed in the venter.

The slender **seta** (plural, setae), as seen in [\[link\]](#), contains tubular cells that transfer nutrients from the base of the sporophyte (the foot) to the

sporangium or **capsule**.



This photograph shows the long slender stems, called setae, connected to capsules of the moss *Thamnobryum alopecurum*. (credit: modification of work by Hermann Schachner)

A structure called a **peristome** increases the spread of spores after the tip of the capsule falls off at dispersal. The concentric tissue around the mouth of the capsule is made of triangular, close-fitting units, a little like “teeth”; these open and close depending on moisture levels, and periodically release spores.

## Section Summary

Seedless nonvascular plants are small, having the gametophyte as the dominant stage of the lifecycle. Without a vascular system and roots, they absorb water and nutrients on all their exposed surfaces. Collectively known as bryophytes, the three main groups include the liverworts, the hornworts, and the mosses. Liverworts are the most primitive plants and are

closely related to the first land plants. Hornworts developed stomata and possess a single chloroplast per cell. Mosses have simple conductive cells and are attached to the substrate by rhizoids. They colonize harsh habitats and can regain moisture after drying out. The moss sporangium is a complex structure that allows release of spores away from the parent plant.

## Art Connections

### Exercise:

#### Problem:

[\[link\]](#) Which of the following statements about the moss life cycle is false?

- a. The mature gametophyte is haploid.
- b. The sporophyte produces haploid spores.
- c. The rhizoid buds to form a mature gametophyte.
- d. The zygote is housed in the venter.

---

#### Solution:

[\[link\]](#) C.

## Review Questions

### Exercise:

#### Problem:

Which of the following structures is not found in bryophytes?

- a. a cellulose cell wall
  - b. chloroplast
  - c. sporangium
  - d. root
-

---

**Solution:**

D

**Exercise:**

**Problem:** Stomata appear in which group of plants?

- a. Charales
- b. liverworts
- c. hornworts
- d. mosses

---

**Solution:**

C

**Exercise:**

**Problem:** The chromosome complement in a moss protonema is:

- a.  $1n$
- b.  $2n$
- c.  $3n$
- d. varies with the size of the protonema

---

**Solution:**

A

**Exercise:**

**Problem:** Why do mosses grow well in the Arctic tundra?

- a. They grow better at cold temperatures.
- b. They do not require moisture.

- c. They do not have true roots and can grow on hard surfaces.
- d. There are no herbivores in the tundra.

---

**Solution:**

C

## Free Response

### Exercise:

**Problem:**

In areas where it rains often, mosses grow on roofs. How do mosses survive on roofs without soil?

---

**Solution:**

Mosses absorb water and nutrients carried by the rain and do not need soil because they do not derive much nutrition from the soil.

### Exercise:

**Problem:** What are the three classes of bryophytes?

---

**Solution:**

The bryophytes are divided into three phyla: the liverworts or Hepaticophyta, the hornworts or Anthocerotophyta, and the mosses or true Bryophyta.

## Glossary

capsule

case of the sporangium in mosses



gemma

(plural, gemmae) leaf fragment that spreads for asexual reproduction

hornworts

group of non-vascular plants in which stomata appear

liverworts

most primitive group of the non-vascular plants

mosses

group of bryophytes in which a primitive conductive system appears

peristome

tissue that surrounds the opening of the capsule and allows periodic release of spores

protonema

tangle of single celled filaments that forms from the haploid spore

rhizoids

thin filaments that anchor the plant to the substrate

seta

stalk that supports the capsule in mosses

## Seedless Vascular Plants

By the end of this section, you will be able to:

- Identify the new traits that first appear in tracheophytes
- Discuss the importance of adaptations to life on land
- Describe the classes of seedless tracheophytes
- Describe the lifecycle of a fern
- Explain the role of seedless vascular plants in the ecosystem

The vascular plants, or **tracheophytes**, are the dominant and most conspicuous group of land plants. More than 260,000 species of tracheophytes represent more than 90 percent of Earth's vegetation. Several evolutionary innovations explain their success and their ability to spread to all habitats.

Bryophytes may have been successful at the transition from an aquatic habitat to land, but they are still dependent on water for reproduction, and absorb moisture and nutrients through the gametophyte surface. The lack of roots for absorbing water and minerals from the soil, as well as a lack of reinforced conducting cells, limits bryophytes to small sizes. Although they may survive in reasonably dry conditions, they cannot reproduce and expand their habitat range in the absence of water. Vascular plants, on the other hand, can achieve enormous heights, thus competing successfully for light. Photosynthetic organs become leaves, and pipe-like cells or vascular tissues transport water, minerals, and fixed carbon throughout the organism.

In seedless vascular plants, the diploid sporophyte is the dominant phase of the lifecycle. The gametophyte is now an inconspicuous, but still independent, organism. Throughout plant evolution, there is an evident reversal of roles in the dominant phase of the lifecycle. Seedless vascular plants still depend on water during fertilization, as the sperm must swim on a layer of moisture to reach the egg. This step in reproduction explains why ferns and their relatives are more abundant in damp environments.

## Vascular Tissue: Xylem and Phloem

The first fossils that show the presence of vascular tissue date to the Silurian period, about 430 million years ago. The simplest arrangement of conductive cells shows a pattern of xylem at the center surrounded by phloem. **Xylem** is the tissue responsible for the storage and long-distance transport of water and nutrients, as well as the transfer of water-soluble growth factors from the organs of synthesis to the target organs. The tissue consists of conducting cells, known as tracheids, and supportive filler tissue, called parenchyma. Xylem conductive cells incorporate the compound **lignin** into their walls, and are thus described as lignified. Lignin itself is a complex polymer that is impermeable to water and confers mechanical strength to vascular tissue. With their rigid cell walls, the xylem cells provide support to the plant and allow it to achieve impressive heights. Tall plants have a selective advantage by being able to reach unfiltered sunlight and disperse their spores or seeds further away, thus expanding their range. By growing higher than other plants, tall trees cast their shadow on shorter plants and limit competition for water and precious nutrients in the soil.

**Phloem** is the second type of vascular tissue; it transports sugars, proteins, and other solutes throughout the plant. Phloem cells are divided into sieve elements (conducting cells) and cells that support the sieve elements. Together, xylem and phloem tissues form the vascular system of plants.

## **Roots: Support for the Plant**

Roots are not well preserved in the fossil record. Nevertheless, it seems that roots appeared later in evolution than vascular tissue. The development of an extensive network of roots represented a significant new feature of vascular plants. Thin rhizoids attached bryophytes to the substrate, but these rather flimsy filaments did not provide a strong anchor for the plant; neither did they absorb substantial amounts of water and nutrients. In contrast, roots, with their prominent vascular tissue system, transfer water and minerals from the soil to the rest of the plant. The extensive network of roots that penetrates deep into the soil to reach sources of water also stabilizes trees by acting as a ballast or anchor. The majority of roots establish a symbiotic relationship with fungi, forming mycorrhizae, which

benefit the plant by greatly increasing the surface area for absorption of water and soil minerals and nutrients.

## Leaves, Sporophylls, and Strobili

A third innovation marks the seedless vascular plants. Accompanying the prominence of the sporophyte and the development of vascular tissue, the appearance of true leaves improved their photosynthetic efficiency. Leaves capture more sunlight with their increased surface area by employing more chloroplasts to trap light energy and convert it to chemical energy, which is then used to fix atmospheric carbon dioxide into carbohydrates. The carbohydrates are exported to the rest of the plant by the conductive cells of phloem tissue.

The existence of two types of morphology suggests that leaves evolved independently in several groups of plants. The first type of leaf is the **microphyll**, or “little leaf,” which can be dated to 350 million years ago in the late Silurian. A microphyll is small and has a simple vascular system. A single unbranched **vein**—a bundle of vascular tissue made of xylem and phloem—runs through the center of the leaf. Microphylls may have originated from the flattening of lateral branches, or from sporangia that lost their reproductive capabilities. Microphylls are present in the club mosses and probably preceded the development of **megaphylls**, or “big leaves”, which are larger leaves with a pattern of branching veins. Megaphylls most likely appeared independently several times during the course of evolution. Their complex networks of veins suggest that several branches may have combined into a flattened organ, with the gaps between the branches being filled with photosynthetic tissue.

In addition to photosynthesis, leaves play another role in the life of the plants. Pine cones, mature fronds of ferns, and flowers are all **sporophylls**—leaves that were modified structurally to bear sporangia. **Strobili** are cone-like structures that contain sporangia. They are prominent in conifers and are commonly known as pine cones.

## Ferns and Other Seedless Vascular Plants

By the late Devonian period, plants had evolved vascular tissue, well-defined leaves, and root systems. With these advantages, plants increased in height and size. During the Carboniferous period, swamp forests of club mosses and horsetails—some specimens reaching heights of more than 30 m (100 ft)—covered most of the land. These forests gave rise to the extensive coal deposits that gave the Carboniferous its name. In seedless vascular plants, the sporophyte became the dominant phase of the lifecycle.

Water is still required for fertilization of seedless vascular plants, and most favor a moist environment. Modern-day seedless tracheophytes include club mosses, horsetails, ferns, and whisk ferns.

### **Phylum Lycopodiophyta: Club Mosses**

The **club mosses**, or phylum **Lycopodiophyta**, are the earliest group of seedless vascular plants. They dominated the landscape of the Carboniferous, growing into tall trees and forming large swamp forests. Today's club mosses are diminutive, evergreen plants consisting of a stem (which may be branched) and microphylls ([\[link\]](#)). The phylum Lycopodiophyta consists of close to 1,200 species, including the quillworts (*Isoetales*), the club mosses (*Lycopodiales*), and spike mosses (*Selaginellales*), none of which are true mosses or bryophytes.

Lycophytes follow the pattern of alternation of generations seen in the bryophytes, except that the sporophyte is the major stage of the lifecycle. The gametophytes do not depend on the sporophyte for nutrients. Some gametophytes develop underground and form mycorrhizal associations with fungi. In club mosses, the sporophyte gives rise to sporophylls arranged in strobili, cone-like structures that give the class its name. Lycophytes can be homosporous or heterosporous.



In the club mosses such as  
*Lycopodium clavatum*,  
sporangia are arranged in  
clusters called strobili. (credit:  
Cory Zanker)

### **Phylum Monilophyta: Class Equisetopsida (Horsetails)**

Horsetails, whisk ferns and ferns belong to the phylum Monilophyta, with **horsetails** placed in the Class Equisetopsida. The single genus *Equisetum* is the survivor of a large group of plants, known as Arthrophyta, which produced large trees and entire swamp forests in the Carboniferous. The plants are usually found in damp environments and marshes ([\[link\]](#)).



Horsetails thrive in a marsh.  
(credit: Myriam Feldman)

The stem of a horsetail is characterized by the presence of joints or nodes, hence the name Arthrophyta (arthro- = "joint"; -phyta = "plant"). Leaves and branches come out as whorls from the evenly spaced joints. The needle-shaped leaves do not contribute greatly to photosynthesis, the majority of which takes place in the green stem ([link](#)).



Thin leaves originating at the joints are noticeable on the horsetail plant.

Horsetails were once used as scrubbing brushes and were nicknamed scouring brushes. (credit: Myriam Feldman)

Silica collects in the epidermal cells, contributing to the stiffness of horsetail plants. Underground stems known as rhizomes anchor the plants to the ground. Modern-day horsetails are homosporous and produce bisexual gametophytes.

### **Phylum Monilophyta: Class Psilotopsida (Whisk Ferns)**

While most ferns form large leaves and branching roots, the **whisk ferns**, Class Psilotopsida, lack both roots and leaves, probably lost by reduction.



Photosynthesis takes place in their green stems, and small yellow knobs form at the tip of the branch stem and contain the sporangia. Whisk ferns were considered an early pterophytes. However, recent comparative DNA analysis suggests that this group may have lost both vascular tissue and roots through evolution, and is more closely related to ferns.



The whisk fern *Psilotum nudum* has conspicuous green stems with knob-shaped sporangia.  
(credit: Forest & Kim Starr)

### **Phylum Monilophyta: Class Psilotopsida (Ferns)**

With their large fronds, **ferns** are the most readily recognizable seedless vascular plants. They are considered the most advanced seedless vascular plants and display characteristics commonly observed in seed plants. More than 20,000 species of ferns live in environments ranging from tropics to temperate forests. Although some species survive in dry environments, most ferns are restricted to moist, shaded places. Ferns made their appearance in the fossil record during the Devonian period and expanded during the Carboniferous.

The dominant stage of the lifecycle of a fern is the sporophyte, which consists of large compound leaves called fronds. Fronds fulfill a double role; they are photosynthetic organs that also carry reproductive organs. The stem may be buried underground as a rhizome, from which adventitious roots grow to absorb water and nutrients from the soil; or, they may grow above ground as a trunk in tree ferns ([\[link\]](#)). **Adventitious** organs are those that grow in unusual places, such as roots growing from the side of a stem.



Some specimens of this short tree-fern species can grow very tall. (credit: Adrian Pingstone)

The tip of a developing fern frond is rolled into a crozier, or fiddlehead ([\[link\]](#)**a** and [\[link\]](#)**b**). Fiddleheads unroll as the frond develops.

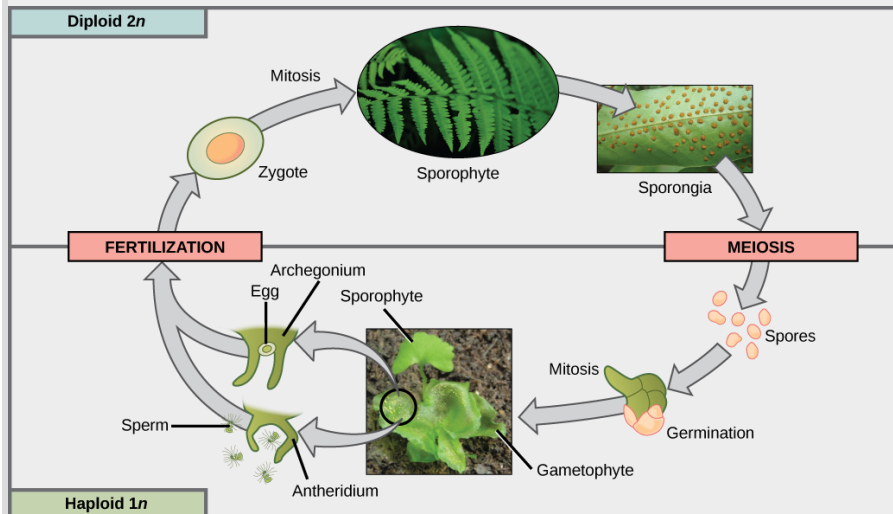


Crozier, or fiddleheads, are the tips of fern fronds. (credit a: modification of work by Cory Zanker; credit b: modification of work by Myriam Feldman)

The lifecycle of a fern is depicted in [\[link\]](#).

**Note:**

**Art Connection**



This life cycle of a fern shows alternation of generations with a dominant sporophyte stage. (credit "fern": modification of work by Cory Zanker; credit "gametophyte": modification of work by "Vlmastra"/Wikimedia Commons)

Which of the following statements about the fern life cycle is false?

- a. Sporangia produce haploid spores.
- b. The sporophyte grows from a gametophyte.

- c. The sporophyte is diploid and the gametophyte is haploid.
- d. Sporangia form on the underside of the gametophyte.

**Note:**

Link to Learning



To see an animation of the lifecycle of a fern and to test your knowledge, go to the [website](#).

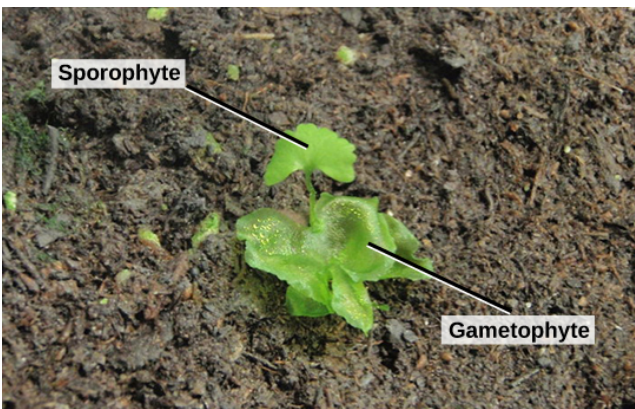
Most ferns produce the same type of spores and are therefore homosporous. The diploid sporophyte is the most conspicuous stage of the lifecycle. On the underside of its mature fronds, sori (singular, sorus) form as small clusters where sporangia develop ([link](#)).



Sori appear as small bumps on the underside of a fern frond. (credit:

Myriam Feldman)

Inside the sori, spores are produced by meiosis and released into the air. Those that land on a suitable substrate germinate and form a heart-shaped gametophyte, which is attached to the ground by thin filamentous rhizoids ([link](#)).



Shown here are a young sporophyte (upper part of image) and a heart-shaped gametophyte (bottom part of image). (credit: modification of work by "Vlmastra"/Wikimedia Commons)

The inconspicuous gametophyte harbors both sex gametangia. Flagellated sperm released from the antheridium swim on a wet surface to the archegonium, where the egg is fertilized. The newly formed zygote grows into a sporophyte that emerges from the gametophyte and grows by mitosis into the next generation sporophyte.

**Note:****Career Connection****Landscape Designer**

Looking at the well-laid parterres of flowers and fountains in the grounds of royal castles and historic houses of Europe, it's clear that the gardens' creators knew about more than art and design. They were also familiar with the biology of the plants they chose. Landscape design also has strong roots in the United States' tradition. A prime example of early American classical design is Monticello: Thomas Jefferson's private estate. Among his many interests, Jefferson maintained a strong passion for botany.

Landscape layout can encompass a small private space, like a backyard garden; public gathering places, like Central Park in New York City; or an entire city plan, like Pierre L'Enfant's design for Washington, DC.

A landscape designer will plan traditional public spaces—such as botanical gardens, parks, college campuses, gardens, and larger developments—as well as natural areas and private gardens. The restoration of natural places encroached on by human intervention, such as wetlands, also requires the expertise of a landscape designer.

With such an array of necessary skills, a landscape designer's education includes a solid background in botany, soil science, plant pathology, entomology, and horticulture. Coursework in architecture and design software is also required for the completion of the degree. The successful design of a landscape rests on an extensive knowledge of plant growth requirements, such as light and shade, moisture levels, compatibility of different species, and susceptibility to pathogens and pests. Mosses and ferns will thrive in a shaded area, where fountains provide moisture; cacti, on the other hand, would not fare well in that environment. The future growth of individual plants must be taken into account, to avoid crowding and competition for light and nutrients. The appearance of the space over time is also of concern. Shapes, colors, and biology must be balanced for a well-maintained and sustainable green space. Art, architecture, and biology blend in a beautifully designed and implemented landscape.





This landscaped border at a college campus was designed by students in the horticulture and landscaping department of the college. (credit: Myriam Feldman)

## The Importance of Seedless Vascular Plants

Mosses and liverworts are often the first macroscopic organisms to colonize an area, both in a primary succession—where bare land is settled for the first time by living organisms—or in a secondary succession, where soil remains intact after a catastrophic event wipes out many existing species. Their spores are carried by the wind, birds, or insects. Once mosses and liverworts are established, they provide food and shelter for other species. In a hostile environment, like the tundra where the soil is frozen, bryophytes grow well because they do not have roots and can dry and rehydrate rapidly once water is again available. Mosses are at the base of the food chain in the tundra biome. Many species—from small insects to musk oxen and reindeer—depend on mosses for food. In turn, predators feed on the herbivores, which are the primary consumers. Some reports indicate that bryophytes make the soil more amenable to colonization by other plants. Because they establish symbiotic relationships with nitrogen-fixing cyanobacteria, mosses replenish the soil with nitrogen.

At the end of the nineteenth century, scientists observed that lichens and mosses were becoming increasingly rare in urban and suburban areas. Since bryophytes have neither a root system for absorption of water and nutrients, nor a cuticle layer that protects them from desiccation, pollutants in rainwater readily penetrate their tissues; they absorb moisture and nutrients through their entire exposed surfaces. Therefore, pollutants dissolved in rainwater penetrate plant tissues readily and have a larger impact on mosses than on other plants. The disappearance of mosses can be considered a bioindicator for the level of pollution in the environment.

Ferns contribute to the environment by promoting the weathering of rock, accelerating the formation of topsoil, and slowing down erosion by spreading rhizomes in the soil. The water ferns of the genus *Azolla* harbor nitrogen-fixing cyanobacteria and restore this important nutrient to aquatic habitats.

Seedless plants have historically played a role in human life through uses as tools, fuel, and medicine. Dried **peat moss**, *Sphagnum*, is commonly used as fuel in some parts of Europe and is considered a renewable resource. *Sphagnum* bogs ([\[link\]](#)) are cultivated with cranberry and blueberry bushes. The ability of *Sphagnum* to hold moisture makes the moss a common soil conditioner. Florists use blocks of *Sphagnum* to maintain moisture for floral arrangements.



*Sphagnum acutifolium* is dried peat moss and can be used as



fuel. (credit: Ken Goulding)

The attractive fronds of ferns make them a favorite ornamental plant. Because they thrive in low light, they are well suited as house plants. More importantly, fiddleheads are a traditional spring food of Native Americans in the Pacific Northwest, and are popular as a side dish in French cuisine. The licorice fern, *Polypodium glycyrrhiza*, is part of the diet of the Pacific Northwest coastal tribes, owing in part to the sweetness of its rhizomes. It has a faint licorice taste and serves as a sweetener. The rhizome also figures in the pharmacopeia of Native Americans for its medicinal properties and is used as a remedy for sore throat.

**Note:**

Link to Learning



Go to this [website](#) to learn how to identify fern species based upon their fiddleheads.

By far the greatest impact of seedless vascular plants on human life, however, comes from their extinct progenitors. The tall club mosses, horsetails, and tree-like ferns that flourished in the swampy forests of the Carboniferous period gave rise to large deposits of coal throughout the world. Coal provided an abundant source of energy during the Industrial Revolution, which had tremendous consequences on human societies, including rapid technological progress and growth of large cities, as well as

the degradation of the environment. Coal is still a prime source of energy and also a major contributor to global warming.

## Section Summary

Vascular systems consist of xylem tissue, which transports water and minerals, and phloem tissue, which transports sugars and proteins. With the development of the vascular system, there appeared leaves to act as large photosynthetic organs, and roots to access water from the ground. Small uncomplicated leaves are microphylls. Large leaves with vein patterns are megaphylls. Modified leaves that bear sporangia are sporophylls. Some sporophylls are arranged in cone structures called strobili.

The seedless vascular plants include club mosses, which are the most primitive; whisk ferns, which lost leaves and roots by reductive evolution; and horsetails and ferns. Ferns are the most advanced group of seedless vascular plants. They are distinguished by large leaves called fronds and small sporangia-containing structures called sori, which are found on the underside of the fronds.

Mosses play an essential role in the balance of the ecosystems; they are pioneering species that colonize bare or devastated environments and make it possible for a succession to occur. They contribute to the enrichment of the soil and provide shelter and nutrients for animals in hostile environments. Mosses and ferns can be used as fuels and serve culinary, medical, and decorative purposes.

## Art Connections

### Exercise:

#### Problem:

[\[link\]](#) Which of the following statements about the fern life cycle is false?

- a. Sporangia produce haploid spores.

- b. The sporophyte grows from a gametophyte.
- c. The sporophyte is diploid and the gametophyte is haploid.
- d. Sporangia form on the underside of the gametophyte.

---

**Solution:**

[\[link\]](#) D.

## Review Questions

**Exercise:**

**Problem:** Microphylls are characteristic of which types of plants?

- a. mosses
- b. liverworts
- c. club mosses
- d. ferns

---

**Solution:**

C

**Exercise:**

**Problem:**

A plant in the understory of a forest displays a segmented stem and slender leaves arranged in a whorl. It is probably a \_\_\_\_\_.

- a. club moss
- b. whisk fern
- c. fern
- d. horsetail

---

**Solution:**

D

**Exercise:**

**Problem:**

The following structures are found on the underside of fern leaves and contain sporangia:

- a. sori
- b. rhizomes
- c. megaphylls
- d. microphylls

---

**Solution:**

A

**Exercise:**

**Problem:** The dominant organism in fern is the \_\_\_\_\_.

- a. sperm
- b. spore
- c. gamete
- d. sporophyte

---

**Solution:**

D

**Exercise:**

**Problem:** What seedless plant is a renewable source of energy?

- a. club moss
- b. horsetail
- c. sphagnum moss
- d. fern

---

**Solution:**

C

**Exercise:**

**Problem:** How do mosses contribute to returning nitrogen to the soil?

- a. Mosses fix nitrogen from the air.
- b. Mosses harbor cyanobacteria that fix nitrogen.
- c. Mosses die and return nitrogen to the soil.
- d. Mosses decompose rocks and release nitrogen.

---

**Solution:**

D

## Free Response

**Exercise:**

**Problem:**

How did the development of a vascular system contribute to the increase in size of plants?

---

**Solution:**

Plants became able to transport water and nutrients and not be limited by rates of diffusion. Vascularization allowed the development of

leaves, which increased efficiency of photosynthesis and provided more energy for plant growth.

**Exercise:**

**Problem:**

Which plant is considered the most advanced seedless vascular plant and why?

---

**Solution:**

Ferns are considered the most advanced seedless vascular plants, because they display characteristics commonly observed in seed plants—they form large leaves and branching roots.

**Glossary**

adventitious

describes an organ that grows in an unusual place, such as a roots growing from the side of a stem

club mosses

earliest group of seedless vascular plants

fern

seedless vascular plant that produces large fronds; the most advanced group of seedless vascular plants

horsetail

seedless vascular plant characterized by joints

lignin

complex polymer impermeable to water

lycophyte

club moss

megaphyll

larger leaves with a pattern of branching veins

microphyll

small size and simple vascular system with a single unbranched vein

peat moss

Sphagnum

phloem

tissue responsible for transport of sugars, proteins, and other solutes

sporophyll

leaf modified structurally to bear sporangia

strobili

cone-like structures that contain the sporangia

tracheophyte

vascular plant

vein

bundle of vascular tissue made of xylem and phloem

whisk fern

seedless vascular plant that lost roots and leaves by reduction

xylem

tissue responsible for long-distance transport of water and nutrients

## Introduction

class="introduction"

Seed plants  
dominate  
the  
landscape  
and play an  
integral role  
in human  
societies. (a)  
Palm trees  
grow along  
the  
shoreline;  
(b) wheat is  
a crop  
grown in  
most of the  
world; (c)  
the flower  
of the cotton  
plant  
produces  
fibers that  
are woven  
into fabric;  
(d) the  
potent  
alkaloids of  
the beautiful  
opium  
poppy have  
influenced  
human life  
both as a  
medicinal



remedy and  
as a  
dangerously  
addictive  
drug. (credit  
a:  
modification  
n of work  
by Ryan  
Kozie;  
credit b:  
modification  
n of work  
by Stephen  
Ausmus;  
credit c:  
modification  
n of work  
by David  
Nance;  
credit d:  
modification  
n of work  
by Jolly  
Janner)



(a)

(b)

(c)

(d)

The lush palms on tropical shorelines do not depend on water for the dispersal of their pollen, fertilization, or the survival of the zygote—unlike mosses, liverworts, and ferns of the terrain. Seed plants, such as palms, have broken free from the need to rely on water for their reproductive needs. They play an integral role in all aspects of life on the planet, shaping the physical terrain, influencing the climate, and maintaining life as we know it. For millennia, human societies have depended on seed plants for nutrition and medicinal compounds: and more recently, for industrial by-products, such as timber and paper, dyes, and textiles. Palms provide materials including rattans, oils, and dates. Wheat is grown to feed both human and animal populations. The fruit of the cotton boll flower is harvested as a boll, with its fibers transformed into clothing or pulp for paper. The showy opium poppy is valued both as an ornamental flower and as a source of potent opiate compounds.

## Evolution of Seed Plants

By the end of this section, you will be able to:

- Explain when seed plants first appeared and when gymnosperms became the dominant plant group
- Describe the two major innovations that allowed seed plants to reproduce in the absence of water
- Discuss the purpose of pollen grains and seeds
- Describe the significance of angiosperms bearing both flowers and fruit

The first plants to colonize land were most likely closely related to modern day mosses (bryophytes) and are thought to have appeared about 500 million years ago. They were followed by liverworts (also bryophytes) and primitive vascular plants—the pterophytes—from which modern ferns are derived. The lifecycle of bryophytes and pterophytes is characterized by the alternation of generations, like gymnosperms and angiosperms; what sets bryophytes and pterophytes apart from gymnosperms and angiosperms is their reproductive requirement for water. The completion of the bryophyte and pterophyte life cycle requires water because the male gametophyte releases sperm, which must swim—propelled by their flagella—to reach and fertilize the female gamete or egg. After fertilization, the zygote matures and grows into a sporophyte, which in turn will form sporangia or "spore vessels." In the sporangia, mother cells undergo meiosis and produce the haploid spores. Release of spores in a suitable environment will lead to germination and a new generation of gametophytes.

In seed plants, the evolutionary trend led to a dominant sporophyte generation, and at the same time, a systematic reduction in the size of the gametophyte: from a conspicuous structure to a microscopic cluster of cells enclosed in the tissues of the sporophyte. Whereas lower vascular plants, such as club mosses and ferns, are mostly homosporous (produce only one type of spore), all seed plants, or **spermatophytes**, are heterosporous. They form two types of spores: megaspores (female) and microspores (male). Megaspores develop into female gametophytes that produce eggs, and microspores mature into male gametophytes that generate sperm. Because the gametophytes mature within the spores, they are not free-living, as are

the gametophytes of other seedless vascular plants. Heterosporous seedless plants are seen as the evolutionary forerunners of seed plants.

Seeds and pollen—two critical adaptations to drought, and to reproduction that doesn't require water—distinguish seed plants from other (seedless) vascular plants. Both adaptations were required for the colonization of land begun by the bryophytes and their ancestors. Fossils place the earliest distinct seed plants at about 350 million years ago. The first reliable record of gymnosperms dates their appearance to the Pennsylvanian period, about 319 million years ago ([link](#)). Gymnosperms were preceded by **progymnosperms**, the first naked seed plants, which arose about 380 million years ago. Progymnosperms were a transitional group of plants that superficially resembled conifers (cone bearers) because they produced wood from the secondary growth of the vascular tissues; however, they still reproduced like ferns, releasing spores into the environment. Gymnosperms dominated the landscape in the early (Triassic) and middle (Jurassic) Mesozoic era. Angiosperms surpassed gymnosperms by the middle of the Cretaceous (about 100 million years ago) in the late Mesozoic era, and today are the most abundant plant group in most terrestrial biomes.

EON	ERA	PERIOD	MILLIONS OF YEARS AGO
Phanerozoic	Cenozoic	Quaternary	1.6
		Tertiary	66
	Mesozoic	Cretaceous	138
		Jurassic	205
		Triassic	240
	Paleozoic	Permian	290
		Pennsylvanian	330
		Mississippian	360
		Devonian	410
		Silurian	435
		Ordovician	500
		Cambrian	570
Proterozoic	Late Proterozoic Middle Proterozoic Early Proterozoic		2500
Archean	Late Archean Middle Archean Early Archean		3800?
Pre-Archean			

Various plant species evolved in different eras. (credit: United States

## Geological Survey)

Pollen and seed were innovative structures that allowed seed plants to break their dependence on water for reproduction and development of the embryo, and to conquer dry land. The **pollen grains** are the male gametophytes, which contain the sperm (gametes) of the plant. The small haploid ( $1n$ ) cells are encased in a protective coat that prevents desiccation (drying out) and mechanical damage. Pollen grains can travel far from their original sporophyte, spreading the plant's genes. The **seed** offers the embryo protection, nourishment, and a mechanism to maintain dormancy for tens or even thousands of years, ensuring germination can occur when growth conditions are optimal. Seeds therefore allow plants to disperse the next generation through both space and time. With such evolutionary advantages, seed plants have become the most successful and familiar group of plants, in part because of their size and striking appearance.

### Evolution of Gymnosperms

The fossil plant *Elkinsia polymorpha*, a "seed fern" from the Devonian period—about 400 million years ago—is considered the earliest seed plant known to date. Seed ferns ([\[link\]](#)) produced their seeds along their branches without specialized structures. What makes them the first true seed plants is that they developed structures called cupules to enclose and protect the **ovule**—the female gametophyte and associated tissues—which develops into a seed upon fertilization. Seed plants resembling modern tree ferns became more numerous and diverse in the coal swamps of the Carboniferous period.



This fossilized leaf is from *Glossopteris*, a seed fern that thrived during the Permian age (290–240 million years ago). (credit: D.L. Schmidt, USGS)

Fossil records indicate the first gymnosperms (progymnosperms) most likely originated in the Paleozoic era, during the middle Devonian period: about 390 million years ago. Following the wet Mississippian and Pennsylvanian periods, which were dominated by giant fern trees, the Permian period was dry. This gave a reproductive edge to seed plants, which are better adapted to survive dry spells. The Ginkgoales, a group of gymnosperms with only one surviving species—the *Ginkgo biloba*—were the first gymnosperms to appear during the lower Jurassic. Gymnosperms expanded in the Mesozoic era (about 240 million years ago), supplanting ferns in the landscape, and reaching their greatest diversity during this time. The Jurassic period was as much the age of the cycads (palm-tree-like gymnosperms) as the age of the dinosaurs. Ginkgoales and the more familiar conifers also dotted the landscape. Although angiosperms (flowering plants) are the major form of plant life in most biomes, gymnosperms still dominate some ecosystems, such as the taiga (boreal forests) and the alpine forests at higher mountain elevations ([\[link\]](#)) because of their adaptation to cold and dry growth conditions.





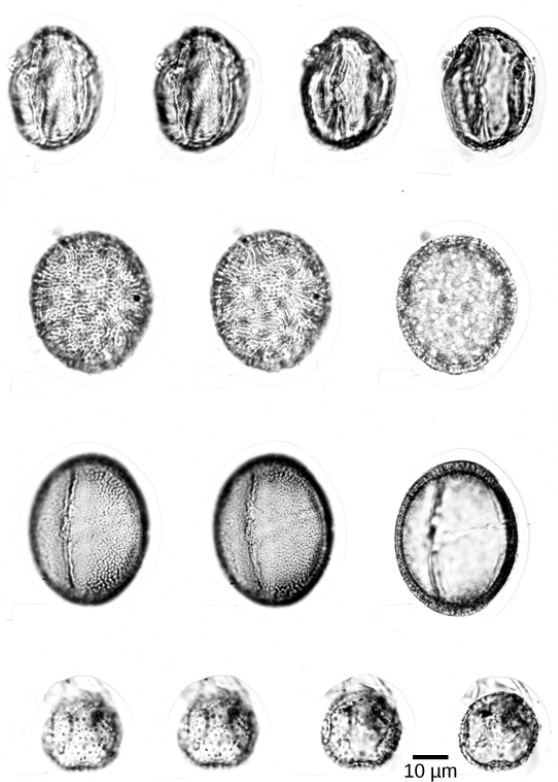
This boreal forest (taiga) has low-lying plants and conifer trees.  
(credit: L.B. Brubaker, NOAA)

## **Seeds and Pollen as an Evolutionary Adaptation to Dry Land**

Unlike bryophyte and fern spores (which are haploid cells dependent on moisture for rapid development of gametophytes), seeds contain a diploid embryo that will germinate into a sporophyte. Storage tissue to sustain growth and a protective coat give seeds their superior evolutionary advantage. Several layers of hardened tissue prevent desiccation, and free reproduction from the need for a constant supply of water. Furthermore, seeds remain in a state of dormancy—induced by desiccation and the hormone abscisic acid—until conditions for growth become favorable. Whether blown by the wind, floating on water, or carried away by animals, seeds are scattered in an expanding geographic range, thus avoiding competition with the parent plant.

Pollen grains ([link](#)) are male gametophytes and are carried by wind, water, or a pollinator. The whole structure is protected from desiccation and can reach the female organs without dependence on water. Male gametes reach female gametophyte and the egg cell gamete through a pollen tube: an extension of a cell within the pollen grain. The sperm of modern

gymnosperms lack flagella, but in cycads and the *Ginkgo*, the sperm still possess flagella that allow them to swim down the **pollen tube** to the female gamete; however, they are enclosed in a pollen grain.



This fossilized pollen is from a Buckbean fen core found in Yellowstone National Park, Wyoming. The pollen is magnified 1,054 times.  
(credit: R.G. Baker, USGS;  
scale-bar data from Matt Russell)

## Evolution of Angiosperms



Undisputed fossil records place the massive appearance and diversification of angiosperms in the middle to late Mesozoic era. Angiosperms (“seed in a vessel”) produce a flower containing male and/or female reproductive structures. Fossil evidence ([\[link\]](#)) indicates that flowering plants first appeared in the Lower Cretaceous, about 125 million years ago, and were rapidly diversifying by the Middle Cretaceous, about 100 million years ago. Earlier traces of angiosperms are scarce. Fossilized pollen recovered from Jurassic geological material has been attributed to angiosperms. A few early Cretaceous rocks show clear imprints of leaves resembling angiosperm leaves. By the mid-Cretaceous, a staggering number of diverse flowering plants crowd the fossil record. The same geological period is also marked by the appearance of many modern groups of insects, including pollinating insects that played a key role in ecology and the evolution of flowering plants.

Although several hypotheses have been offered to explain this sudden profusion and variety of flowering plants, none have garnered the consensus of paleobotanists (scientists who study ancient plants). New data in comparative genomics and paleobotany have, however, shed some light on the evolution of angiosperms. Rather than being derived from gymnosperms, angiosperms form a sister clade (a species and its descendants) that developed in parallel with the gymnosperms. The two innovative structures of flowers and fruit represent an improved reproductive strategy that served to protect the embryo, while increasing genetic variability and range. Paleobotanists debate whether angiosperms evolved from small woody bushes, or were basal angiosperms related to tropical grasses. Both views draw support from cladistics studies, and the so-called woody magnoliid hypothesis—which proposes that the early ancestors of angiosperms were shrubs—also offers molecular biological evidence.

The most primitive living angiosperm is considered to be *Amborella trichopoda*, a small plant native to the rainforest of New Caledonia, an island in the South Pacific. Analysis of the genome of *A. trichopoda* has shown that it is related to all existing flowering plants and belongs to the oldest confirmed branch of the angiosperm family tree. A few other angiosperm groups called basal angiosperms, are viewed as primitive

because they branched off early from the phylogenetic tree. Most modern angiosperms are classified as either monocots or eudicots, based on the structure of their leaves and embryos. Basal angiosperms, such as water lilies, are considered more primitive because they share morphological traits with both monocots and eudicots.



This leaf imprint shows  
a *Ficus speciosissima*,  
an angiosperm that  
flourished during the  
Cretaceous period.  
(credit: W. T. Lee,  
USGS)

## **Flowers and Fruits as an Evolutionary Adaptation**

Angiosperms produce their gametes in separate organs, which are usually housed in a **flower**. Both fertilization and embryo development take place inside an anatomical structure that provides a stable system of sexual reproduction largely sheltered from environmental fluctuations. Flowering plants are the most diverse phylum on Earth after insects; flowers come in a bewildering array of sizes, shapes, colors, smells, and arrangements. Most flowers have a mutualistic pollinator, with the distinctive features of flowers reflecting the nature of the pollination agent. The relationship between pollinator and flower characteristics is one of the great examples of coevolution.

Following fertilization of the egg, the ovule grows into a seed. The surrounding tissues of the ovary thicken, developing into a **fruit** that will protect the seed and often ensure its dispersal over a wide geographic range. Not all fruits develop from an ovary; such structures are “false fruits.” Like flowers, fruit can vary tremendously in appearance, size, smell, and taste. Tomatoes, walnut shells and avocados are all examples of fruit. As with pollen and seeds, fruits also act as agents of dispersal. Some may be carried away by the wind. Many attract animals that will eat the fruit and pass the seeds through their digestive systems, then deposit the seeds in another location. Cockleburs are covered with stiff, hooked spines that can hook into fur (or clothing) and hitch a ride on an animal for long distances. The cockleburs that clung to the velvet trousers of an enterprising Swiss hiker, George de Mestral, inspired his invention of the loop and hook fastener he named Velcro.

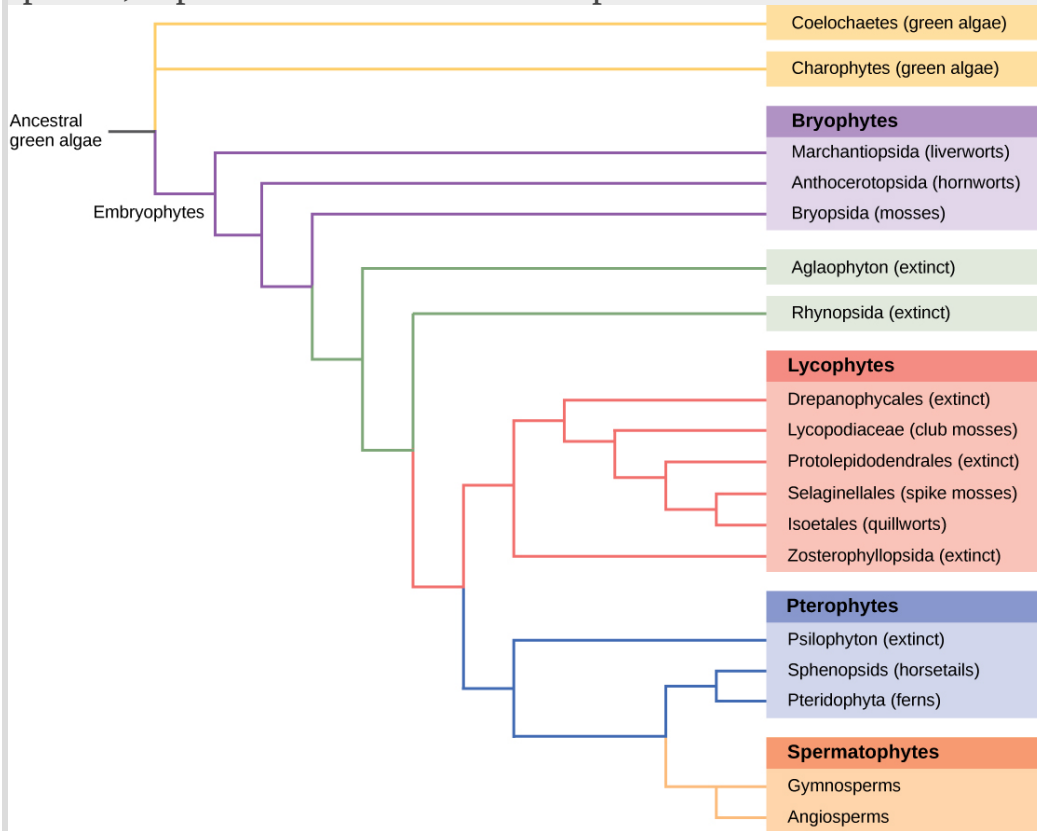
**Note:**

Evolution Connection

**Building Phylogenetic Trees with Analysis of DNA Sequence****Alignments**

All living organisms display patterns of relationships derived from their evolutionary history. Phylogeny is the science that describes the relative connections between organisms, in terms of ancestral and descendant species. Phylogenetic trees, such as the plant evolutionary history shown in [\[link\]](#), are tree-like branching diagrams that depict these relationships.

Species are found at the tips of the branches. Each branching point, called a node, is the point at which a single taxonomic group (taxon), such as a species, separates into two or more species.



This phylogenetic tree shows the evolutionary relationships of plants.

Phylogenetic trees have been built to describe the relationships between species since Darwin's time. Traditional methods involve comparison of homologous anatomical structures and embryonic development, assuming that closely related organisms share anatomical features during embryo development. Some traits that disappear in the adult are present in the embryo; for example, a human fetus, at one point, has a tail. The study of fossil records shows the intermediate stages that link an ancestral form to its descendants. Most of these approaches are imprecise and lend themselves to multiple interpretations. As the tools of molecular biology and computational analysis have been developed and perfected in recent years, a new generation of tree-building methods has taken shape. The key

assumption is that genes for essential proteins or RNA structures, such as the ribosomal RNA, are inherently conserved because mutations (changes in the DNA sequence) could compromise the survival of the organism. DNA from minute amounts of living organisms or fossils can be amplified by polymerase chain reaction (PCR) and sequenced, targeting the regions of the genome that are most likely to be conserved between species. The genes encoding the ribosomal RNA from the small 18S subunit and plastid genes are frequently chosen for DNA alignment analysis. Once the sequences of interest are obtained, they are compared with existing sequences in databases such as GenBank, which is maintained by The National Center for Biotechnology Information. A number of computational tools are available to align and analyze sequences. Sophisticated computer analysis programs determine the percentage of sequence identity or homology. Sequence homology can be used to estimate the evolutionary distance between two DNA sequences and reflect the time elapsed since the genes separated from a common ancestor. Molecular analysis has revolutionized phylogenetic trees. In some cases, prior results from morphological studies have been confirmed: for example, confirming *Amborella trichopoda* as the most primitive angiosperm known. However, some groups and relationships have been rearranged as a result of DNA analysis.

## Section Summary

Seed plants appeared about one million years ago, during the Carboniferous period. Two major innovations—seed and pollen—allowed seed plants to reproduce in the absence of water. The gametophytes of seed plants shrank, while the sporophytes became prominent structures and the diploid stage became the longest phase of the lifecycle. Gymnosperms became the dominant group during the Triassic. In these, pollen grains and seeds protect against desiccation. The seed, unlike a spore, is a diploid embryo surrounded by storage tissue and protective layers. It is equipped to delay germination until growth conditions are optimal. Angiosperms bear both flowers and fruit. The structures protect the gametes and the embryo during

its development. Angiosperms appeared during the Mesozoic era and have become the dominant plant life in terrestrial habitats.

## Review Questions

### Exercise:

**Problem:** Seed plants are \_\_\_\_\_.

- a. all homosporous.
- b. mostly homosporous with some heterosporous.
- c. mostly heterosporous with some homosporous.
- d. all heterosporous.

---

### Solution:

D

### Exercise:

**Problem:**

Besides the seed, what other major structure diminishes a plant's reliance on water for reproduction?

- a. flower
- b. fruit
- c. pollen
- d. spore

---

### Solution:

A

### Exercise:

**Problem:**

In which of the following geological periods would gymnosperms dominate the landscape?

- a. Carboniferous
- b. Permian
- c. Triassic
- d. Eocene (present)

---

**Solution:**

C

**Exercise:****Problem:**

Which of the following structures widens the geographic range of a species and is an agent of dispersal?

- a. seed
- b. flower
- c. leaf
- d. root

---

**Solution:**

A

**Free Response****Exercise:**

**Problem:**

The Triassic Period was marked by the increase in number and variety of angiosperms. Insects also diversified enormously during the same period. Can you propose the reason or reasons that could foster coevolution?

---

**Solution:**

Both pollination and herbivory contributed to diversity, with plants needing to attract some insects and repel others.

**Exercise:****Problem:**

What role did the adaptations of seed and pollen play in the development and expansion of seed plants?

---

**Solution:**

Seeds and pollen allowed plants to reproduce in absence of water. This allowed them to expand their range onto dry land and to survive drought conditions.

**Glossary****flower**

branches specialized for reproduction found in some seed-bearing plants, containing either specialized male or female organs or both male and female organs

**fruit**

thickened tissue derived from ovary wall that protects the embryo after fertilization and facilitates seed dispersal

**ovule**

female gametophyte



pollen grain

structure containing the male gametophyte of the plant

pollen tube

extension from the pollen grain that delivers sperm to the egg cell

progymnosperm

transitional group of plants that resembled conifers because they produced wood, yet still reproduced like ferns

seed

structure containing the embryo, storage tissue and protective coat

spermatophyte

seed plant; from the Greek *sperm* (seed) and *phyte* (plant)

## Gymnosperms

By the end of this section, you will be able to:

- Discuss the type of seeds produced by gymnosperms, as well as other characteristics of gymnosperms
- State which period saw the first appearance of gymnosperms and explain when they were the dominant plant life
- List the four groups of modern-day gymnosperms and provide examples of each

**Gymnosperms**, meaning “naked seeds,” are a diverse group of seed plants and are paraphyletic. Paraphyletic groups are those in which not all members are descendants of a single common ancestor. Their characteristics include naked seeds, separate female and male gametes, pollination by wind, and tracheids (which transport water and solutes in the vascular system).

Gymnosperm seeds are not enclosed in an ovary; rather, they are exposed on cones or modified leaves. Sporophylls are specialized leaves that produce sporangia. The term **strobilus** (plural = strobili) describes a tight arrangement of sporophylls around a central stalk, as seen in cones. Some seeds are enveloped by sporophyte tissues upon maturation. The layer of sporophyte tissue that surrounds the megasporangium, and later, the embryo, is called the **integument**.

Gymnosperms were the dominant phylum in Mesozoic era. They are adapted to live where fresh water is scarce during part of the year, or in the nitrogen-poor soil of a bog. Therefore, they are still the prominent phylum in the coniferous biome or taiga, where the evergreen conifers have a selective advantage in cold and dry weather. Evergreen conifers continue low levels of photosynthesis during the cold months, and are ready to take advantage of the first sunny days of spring. One disadvantage is that conifers are more susceptible than deciduous trees to infestations because conifers do not lose their leaves all at once. They cannot, therefore, shed parasites and restart with a fresh supply of leaves in spring.

The life cycle of a gymnosperm involves alternation of generations, with a dominant sporophyte in which the female gametophyte resides, and reduced

gametophytes. All gymnosperms are heterosporous. The male and female reproductive organs can form in cones or strobili. Male and female sporangia are produced either on the same plant, described as **monoecious** (“one home” or bisexual), or on separate plants, referred to as **dioecious** (“two homes” or unisexual) plants. The life cycle of a conifer will serve as our example of reproduction in gymnosperms.

## Life Cycle of a Conifer

Pine trees are conifers (cone bearing) and carry both male and female sporophylls on the same mature sporophyte. Therefore, they are monoecious plants. Like all gymnosperms, pines are heterosporous and generate two different types of spores: male microspores and female megaspores. In the male cones, or staminate cones, the **microsporocytes** give rise to pollen grains by meiosis. In the spring, large amounts of yellow pollen are released and carried by the wind. Some gametophytes will land on a female cone. Pollination is defined as the initiation of pollen tube growth. The pollen tube develops slowly, and the generative cell in the pollen grain divides into two haploid sperm cells by mitosis. At fertilization, one of the sperm cells will finally unite its haploid nucleus with the haploid nucleus of a haploid egg cell.

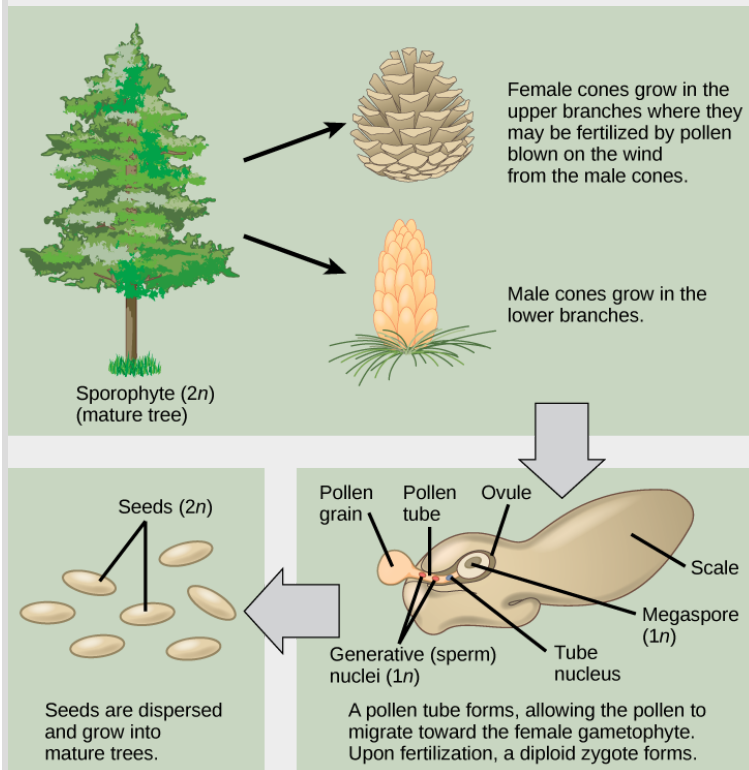
Female cones, or **ovulate cones**, contain two ovules per scale. One megaspore mother cell, or **megasporocyte**, undergoes meiosis in each ovule. Three of the four cells break down; only a single surviving cell will develop into a female multicellular gametophyte, which encloses archegonia (an archegonium is a reproductive organ that contains a single large egg). Upon fertilization, the diploid egg will give rise to the embryo, which is enclosed in a seed coat of tissue from the parent plant. Fertilization and seed development is a long process in pine trees: it may take up to two years after pollination. The seed that is formed contains three generations of tissues: the seed coat that originates from the sporophyte tissue, the gametophyte that will provide nutrients, and the embryo itself.

[\[link\]](#) illustrates the life cycle of a conifer. The sporophyte ( $2n$ ) phase is the longest phase in the life of a gymnosperm. The gametophytes ( $1n$ )—microspores and megaspores—are reduced in size. It may take more than

year between pollination and fertilization while the pollen tube grows towards the megasporocyte ( $2n$ ), which undergoes meiosis into megaspores. The megaspores will mature into eggs ( $1n$ ).

**Note:**

**Art Connection**



This image shows the life cycle of a conifer. Pollen from male cones blows up into upper branches, where it fertilizes female cones.

At what stage does the diploid zygote form?

- a. when the female cone begins to bud from the tree
- b. at fertilization
- c. when the seeds drop from the tree

d. when the pollen tube begins to grow

**Note:**

Link to Learning



Watch this video to see the process of seed production in gymnosperms.  
<https://www.openstaxcollege.org/l/gymnosperm2>

## Diversity of Gymnosperms

Modern gymnosperms are classified into four phyla. Coniferophyta, Cycadophyta, and Ginkgophyta are similar in their production of secondary cambium (cells that generate the vascular system of the trunk or stem and are partially specialized for water transportation) and their pattern of seed development. However, the three phyla are not closely related phylogenetically to each other. Gnetophyta are considered the closest group to angiosperms because they produce true xylem tissue.

### Conifers (Coniferophyta)

**Conifers** are the dominant phylum of gymnosperms, with the most variety of species ([link](#)). Most are typically tall trees that usually bear scale-like or needle-like leaves. Water evaporation from leaves is reduced by their thin shape and the thick cuticle. Snow slides easily off needle-shaped leaves, keeping the load light and decreasing breaking of branches. Adaptations to

cold and dry weather explain the predominance of conifers at high altitudes and in cold climates. Conifers include familiar evergreen trees such as pines, spruces, firs, cedars, sequoias, and yews. A few species are deciduous and lose their leaves in fall. The European larch and the tamarack are examples of deciduous conifers ([link](#)c). Many coniferous trees are harvested for paper pulp and timber. The wood of conifers is more primitive than the wood of angiosperms; it contains tracheids, but no vessel elements, and is therefore referred to as “soft wood.”



(a)



(b)



(c)



(d)

Conifers are the dominant form of vegetation in cold or arid environments and at high altitudes. Shown here are the (a) evergreen spruce *Picea* sp., (b) juniper *Juniperus* sp., (c) sequoia *Sequoia Semervirens*, which is a deciduous gymnosperm, and (d) the tamarack *Larix larcinia*. Notice the yellow leaves of the tamarack. (credit a: modification of work by

Rosendahl; credit b: modification of work by Alan Levine; credit c: modification of work by Wendy McCormic; credit d: modification of work by Micky Zlimen)

## Cycads

**Cycads** thrive in mild climates, and are often mistaken for palms because of the shape of their large, compound leaves. Cycads bear large cones ([link](#)), and may be pollinated by beetles rather than wind: unusual for a gymnosperm. They dominated the landscape during the age of dinosaurs in the Mesozoic, but only a hundred or so species persisted to modern times. They face possible extinction, and several species are protected through international conventions. Because of their attractive shape, they are often used as ornamental plants in gardens in the tropics and subtropics.



This *Encephalartos ferox* cycad has large cones and broad, fern-like leaves. (credit: Wendy Cutler)



## Ginkgophytes

The single surviving species of the **ginkgophytes** group is the *Ginkgo biloba* ([link](#)). Its fan-shaped leaves—unique among seed plants because they feature a dichotomous venation pattern—turn yellow in autumn and fall from the tree. For centuries, *G. biloba* was cultivated by Chinese Buddhist monks in monasteries, which ensured its preservation. It is planted in public spaces because it is unusually resistant to pollution. Male and female organs are produced on separate plants. Typically, gardeners plant only male trees because the seeds produced by the female plant have an off-putting smell of rancid butter.



This plate from the 1870 book *Flora Japonica, Sectio Prima (Tafelband)* depicts the



leaves and fruit of *Ginkgo biloba*, as drawn by Philipp Franz von Siebold and Joseph Gerhard Zuccarini.

## Gnetophytes

**Gnetophytes** are the closest relative to modern angiosperms, and include three dissimilar genera of plants: *Ephedra*, *Gnetum*, and *Welwitschia* ([link](#)). Like angiosperms, they have broad leaves. In tropical and subtropical zones, gnetophytes are vines or small shrubs. *Ephedra* occurs in dry areas of the West Coast of the United States and Mexico. *Ephedra*'s small, scale-like leaves are the source of the compound ephedrine, which is used in medicine as a potent decongestant. Because ephedrine is similar to amphetamines, both in chemical structure and neurological effects, its use is restricted to prescription drugs. Like angiosperms, but unlike other gymnosperms, all gnetophytes possess vessel elements in their xylem.



(a) *Ephedra*

(b) *Gnetum*

(c) *Welwitschia*

(a) *Ephedra viridis*, known by the common name *Mormon tea*, grows on the West Coast of the United States and Mexico. (b) *Gnetum gnemon* grows in Malaysia. (c) The large *Welwitschia mirabilis* can be found in the Namibian desert. (credit a: modification of work by USDA; credit b: modification of work

by Malcolm Manners; credit c: modification of work by Derek Keats)

**Note:**

Link to Learning



Watch this BBC video describing the amazing strangeness of Welwitschia.  
<https://www.openstaxcollege.org/l/welwitschia2>

## Section Summary

Gymnosperms are heterosporous seed plants that produce naked seeds. They appeared in the Paleozoic period and were the dominant plant life during the Mesozoic. Modern-day gymnosperms belong to four phyla. The largest phylum, Coniferophyta, is represented by conifers, the predominant plants at high altitude and latitude. Cycads (phylum Cycadophyta) resemble palm trees and grow in tropical climates. *Ginkgo biloba* is the only representative of the phylum Ginkgophyta. The last phylum, Gnetophyta, is a diverse group of shrubs that produce vessel elements in their wood.

## Art Connections

### Exercise:

**Problem:** [\[link\]](#) At what stage does the diploid zygote form?

- a. When the female cone begins to bud from the tree
  - b. At fertilization
  - c. When the seeds drop from the tree
  - d. When the pollen tube begins to grow
- 

**Solution:**

[\[link\]](#) B. The diploid zygote forms after the pollen tube has finished forming, so that the male generative nuclei can fuse with the female gametophyte.

## Review Questions

**Exercise:**

**Problem:** Which of the following traits characterizes gymnosperms?

- a. The plants carry exposed seeds on modified leaves.
  - b. Reproductive structures are located in a flower.
  - c. After fertilization, the ovary thickens and forms a fruit.
  - d. The gametophyte is longest phase of the life cycle.
- 

**Solution:**

A

**Exercise:**

**Problem:**

Megasporocytes will eventually produce which of the following?

- a. pollen grain
- b. sporophytes
- c. male gametophytes
- d. female gametophytes

---

**Solution:**

D

**Exercise:**

**Problem:**

What is the ploidy of the following structures: gametophyte, seed, spore, sporophyte?

- a.  $1n$ ,  $1n$ ,  $2n$ ,  $2n$
- b.  $1n$ ,  $2n$ ,  $1n$ ,  $2n$
- c.  $2n$ ,  $1n$ ,  $2n$ ,  $1n$
- d.  $2n$ ,  $2n$ ,  $1n$ ,  $1n$

---

**Solution:**

B

**Exercise:**

**Problem:** In the northern forests of Siberia, a tall tree is most likely a:

- a. conifer
- b. cycad
- c. *Ginkgo biloba*
- d. gnetophyte

---

**Solution:**

A

**Free Response**

**Exercise:****Problem:**

The Mediterranean landscape along the sea shore is dotted with pines and cypresses. The weather is not cold, and the trees grow at sea level. What evolutionary adaptation of conifers makes them suitable to the Mediterranean climate?

---

**Solution:**

The trees are adapted to arid weather, and do not lose as much water due to transpiration as non-conifers.

**Exercise:**

**Problem:** What are the four modern-day phyla of gymnosperms?

---

**Solution:**

The four modern-day phyla of gymnosperms are Coniferophyta, Cycadophyta, Ginkgophyta, and Gnetophyta.

**Glossary**

conifer

dominant phylum of gymnosperms with the most variety of trees

cycad

gymnosperm that grows in tropical climates and resembles a palm tree;  
member of the phylum Cycadophyta

dioecious

describes a species in which the male and female reproductive organs  
are carried on separate specimens

ginkgophyte

gymnosperm with one extant species, the *Ginkgo biloba*: a tree with fan-shaped leaves

gnetophyte

gymnosperm shrub with varied morphological features that produces vessel elements in its woody tissues; the phylum includes the genera *Ephedra*, *Gnetum* and *Welwitschia*

gymnosperm

seed plant with naked seeds (seeds exposed on modified leaves or in cones)

integument

layer of sporophyte tissue that surrounds the megasporangium, and later, the embryo

megasporocyte

megaspore mother cell; larger spore that germinates into a female gametophyte in a heterosporous plant

microsporocyte

smaller spore that produces a male gametophyte in a heterosporous plant

monoecious

describes a species in which the male and female reproductive organs are on the same plant

ovulate cone

cone containing two ovules per scale

strobilus

plant structure with a tight arrangement of sporophylls around a central stalk, as seen in cones or flowers; the male strobilus produces pollen, and the female strobilus produces eggs

## Angiosperms

By the end of this section, you will be able to:

- Explain why angiosperms are the dominant form of plant life in most terrestrial ecosystems
- Describe the main parts of a flower and their purpose
- Detail the life cycle of an angiosperm
- Discuss the two main groups of flowering plants

From their humble and still obscure beginning during the early Jurassic period, the angiosperms—or flowering plants—have evolved to dominate most terrestrial ecosystems ([\[link\]](#)). With more than 250,000 species, the angiosperm phylum (Anthophyta) is second only to insects in terms of diversification.



These flowers grow in a botanical garden border in Bellevue, WA. Flowering plants dominate terrestrial landscapes. The vivid colors of flowers are an adaptation to pollination by animals such as insects and birds. (credit: Myriam Feldman)

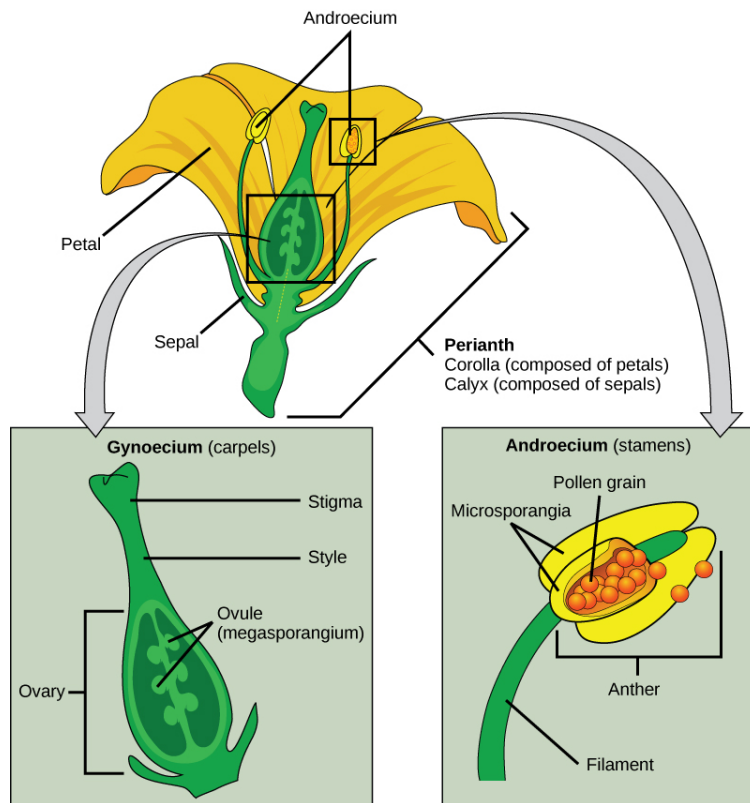
The success of angiosperms is due to two novel reproductive structures: flowers and fruit. The function of the flower is to ensure pollination. Flowers also provide protection for the ovule and developing embryo inside a receptacle. The function of the fruit is seed dispersal. They also protect the developing seed. Different fruit structures or tissues on fruit—such as sweet flesh, wings, parachutes, or spines that grab—reflect the dispersal strategies that help spread seeds.

## Flowers

Flowers are modified leaves, or sporophylls, organized around a central stalk. Although they vary greatly in appearance, all flowers contain the same structures: sepals, petals, carpels, and stamens. The peduncle attaches the flower to the plant. A whorl of **sepals** (collectively called the **calyx**) is located at the base of the peduncle and encloses the unopened floral bud. Sepals are usually photosynthetic organs, although there are some exceptions. For example, the corolla in lilies and tulips consists of three sepals and three petals that look virtually identical. **Petals**, collectively the **corolla**, are located inside the whorl of sepals and often display vivid colors to attract pollinators. Flowers pollinated by wind are usually small, feathery, and visually inconspicuous. Sepals and petals together form the **perianth**. The sexual organs (carpels and stamens) are located at the center of the flower.

As illustrated in [\[link\]](#), styles, stigmas, and ovules constitute the female organ: the **gynoecium** or **carpel**. Flower structure is very diverse, and carpels may be singular, multiple, or fused. Multiple fused carpels comprise a **pistil**. The megaspores and the female gametophytes are produced and protected by the thick tissues of the carpel. A long, thin structure called a **style** leads from the sticky **stigma**, where pollen is deposited, to the **ovary**, enclosed in the carpel. The ovary houses one or more ovules, each of which will develop into a seed upon fertilization. The male reproductive organs, the **stamens** (collectively called the androecium), surround the central carpel. Stamens are composed of a thin stalk called a **filament** and a sac-like structure called the anther. The filament supports the **anther**, where the microspores are produced by meiosis and develop into pollen grains.





This image depicts the structure of a perfect flower. Perfect flowers produce both male and female floral organs. The flower shown has only one carpel, but some flowers have a cluster of carpels.

Together, all the carpels make up the gynoecium. (credit: modification of work by Mariana Ruiz Villareal)

## Fruit

As the seed develops, the walls of the ovary thicken and form the fruit. The seed forms in an ovary, which also enlarges as the seeds grow. In botany, a fertilized and fully grown, ripened ovary is a fruit. Many foods commonly called vegetables are actually fruit. Eggplants, zucchini, string beans, and bell peppers are all technically fruit because they contain seeds and are

derived from the thick ovary tissue. Acorns are nuts, and winged maple whirligigs (whose botanical name is samara) are also fruit. Botanists classify fruit into more than two dozen different categories, only a few of which are actually fleshy and sweet.

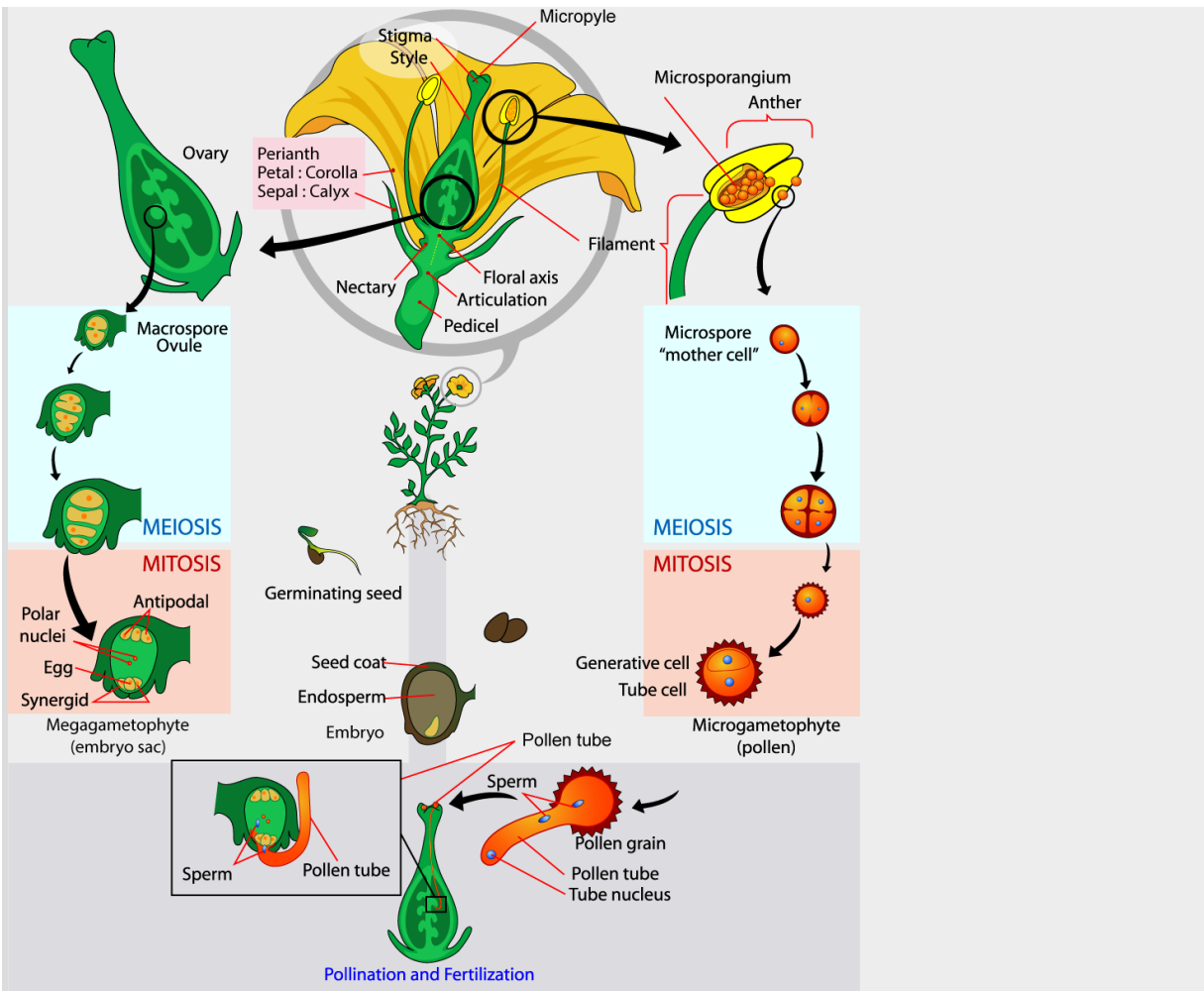
Mature fruit can be fleshy or dry. Fleshy fruit include the familiar berries, peaches, apples, grapes, and tomatoes. Rice, wheat, and nuts are examples of dry fruit. Another distinction is that not all fruits are derived from the ovary. For instance, strawberries are derived from the receptacle and apples from the pericarp, or hypanthium. Some fruits are derived from separate ovaries in a single flower, such as the raspberry. Other fruits, such as the pineapple, form from clusters of flowers. Additionally, some fruits, like watermelon and orange, have rinds. Regardless of how they are formed, fruits are an agent of seed dispersal. The variety of shapes and characteristics reflect the mode of dispersal. Wind carries the light dry fruit of trees and dandelions. Water transports floating coconuts. Some fruits attract herbivores with color or perfume, or as food. Once eaten, tough, undigested seeds are dispersed through the herbivore's feces. Other fruits have burs and hooks to cling to fur and hitch rides on animals.

## **The Life Cycle of an Angiosperm**

The adult, or sporophyte, phase is the main phase of an angiosperm's life cycle ([link](#)). Like gymnosperms, angiosperms are heterosporous. Therefore, they generate microspores, which will generate pollen grains as the male gametophytes, and megaspores, which will form an ovule that contains female gametophytes. Inside the anthers' microsporangia, male gametophytes divide by meiosis to generate haploid microspores, which, in turn, undergo mitosis and give rise to pollen grains. Each pollen grain contains two cells: one generative cell that will divide into two sperm and a second cell that will become the pollen tube cell.

### **Note:**

Art Connection



The life cycle of an angiosperm is shown. Anthers and carpels are structures that shelter the actual gametophytes: the pollen grain and embryo sac. Double fertilization is a process unique to angiosperms. (credit: modification of work by Mariana Ruiz Villareal)

If a flower lacked a megasporangium, what type of gamete would not form? If the flower lacked a microsporangium, what type of gamete would not form?

The ovule, sheltered within the ovary of the carpel, contains the megasporangium protected by two layers of integuments and the ovary wall. Within each megasporangium, a megasporocyte undergoes meiosis, generating four megaspores—three small and one large. Only the large megaspore survives; it produces the female gametophyte, referred to as the embryo sac. The megaspore divides three times to form an eight-cell stage. Four of these cells migrate to each pole of the embryo sac; two come to the equator, and will eventually fuse to form a  $2n$  polar nucleus; the three cells away from the egg form antipodals, and the two cells closest to the egg become the synergids.

The mature embryo sac contains one egg cell, two synergids or “helper” cells, three antipodal cells, and two polar nuclei in a central cell. When a pollen grain reaches the stigma, a pollen tube extends from the grain, grows down the style, and enters through the micropyle: an opening in the integuments of the ovule. The two sperm cells are deposited in the embryo sac.

A double fertilization event then occurs. One sperm and the egg combine, forming a diploid zygote—the future embryo. The other sperm fuses with the  $2n$  polar nuclei, forming a triploid cell that will develop into the endosperm, which is tissue that serves as a food reserve. The zygote develops into an embryo with a radicle, or small root, and one (monocot) or two (dicot) leaf-like organs called **cotyledons**. This difference in the number of embryonic leaves is the basis for the two major groups of angiosperms: the monocots and the eudicots. Seed food reserves are stored outside the embryo, in the form of complex carbohydrates, lipids or proteins. The cotyledons serve as conduits to transmit the broken-down food reserves from their storage site inside the seed to the developing embryo. The seed consists of a toughened layer of integuments forming the coat, the endosperm with food reserves, and at the center, the well-protected embryo.

Most flowers are monoecious or bisexual, which means that they carry both stamens and carpels; only a few species self-pollinate. Monoecious flowers are also known as “perfect” flowers because they contain both types of sex organs ([\[link\]](#)). Both anatomical and environmental barriers promote cross-

pollination mediated by a physical agent (wind or water), or an animal, such as an insect or bird. Cross-pollination increases genetic diversity in a species.

## Diversity of Angiosperms

Angiosperms are classified in a single phylum: the **Anthophyta**. Modern angiosperms appear to be a monophyletic group, which means that they originate from a single ancestor. Flowering plants are divided into two major groups, according to the structure of the cotyledons, pollen grains, and other structures. **Monocots** include grasses and lilies, and eudicots or **dicots** form a polyphyletic group. **Basal angiosperms** are a group of plants that are believed to have branched off before the separation into monocots and eudicots because they exhibit traits from both groups. They are categorized separately in many classification schemes. The *Magnoliidae* (magnolia trees, laurels, and water lilies) and the *Piperaceae* (peppers) belong to the basal angiosperm group.

### Basal Angiosperms

The *Magnoliidae* are represented by the magnolias: tall trees bearing large, fragrant flowers that have many parts and are considered archaic ([\[link\]](#)**d**). Laurel trees produce fragrant leaves and small, inconspicuous flowers. The *Laurales* grow mostly in warmer climates and are small trees and shrubs. Familiar plants in this group include the bay laurel, cinnamon, spice bush ([\[link\]](#)**a**), and avocado tree. The *Nymphaeales* are comprised of the water lilies, lotus ([\[link\]](#)**c**), and similar plants; all species thrive in freshwater biomes, and have leaves that float on the water surface or grow underwater. Water lilies are particularly prized by gardeners, and have graced ponds and pools for thousands of years. The *Piperales* are a group of herbs, shrubs, and small trees that grow in the tropical climates. They have small flowers without petals that are tightly arranged in long spikes. Many species are the source of prized fragrance or spices, for example the berries of *Piper nigrum* ([\[link\]](#)**b**) are the familiar black peppercorns that are used to flavor many dishes.



(a)



(b)



(c)



(d)

The (a) common spicebush belongs to the *Laurales*, the same family as cinnamon and bay laurel. The fruit of (b) the *Piper nigrum* plant is black pepper, the main product that was traded along spice routes. Notice the small, unobtrusive, clustered flowers. (c) Lotus flowers, *Nelumbo nucifera*, have been cultivated since ancient times for their ornamental value; the root of the lotus flower is eaten as a vegetable. The red seeds of (d) a magnolia tree, characteristic of the final stage, are just starting to appear.  
 (credit a: modification of work by Cory Zanker; credit b: modification of work by Franz Eugen Köhler; credit c: modification of work by "berduchwal"/Flickr; credit d:

modification of work by  
"Coastside2"/Wikimedia Commons).

## **Monocots**

Plants in the monocot group are primarily identified as such by the presence of a single cotyledon in the seedling. Other anatomical features shared by monocots include veins that run parallel to the length of the leaves, and flower parts that are arranged in a three- or six-fold symmetry. True woody tissue is rarely found in monocots. In palm trees, vascular and parenchyma tissues produced by the primary and secondary thickening meristems form the trunk. The pollen from the first angiosperms was monosulcate, containing a single furrow or pore through the outer layer. This feature is still seen in the modern monocots. Vascular tissue of the stem is not arranged in any particular pattern. The root system is mostly adventitious and unusually positioned, with no major tap root. The monocots include familiar plants such as the true lilies (which are at the origin of their alternate name of Liliopsida), orchids, grasses, and palms. Many important crops are monocots, such as rice and other cereals, corn, sugar cane, and tropical fruits like bananas and pineapples ([\[link\]](#)).



#### Monocots



(a) Rice



(b) Wheat



(c) Bananas

#### Dicots



(d) Cabbage



(e) Beans



(f) Peaches

The world's major crops are flowering plants. (a) Rice, (b) wheat, and (c) bananas are monocots, while (d) cabbage, (e) beans, and (f) peaches are dicots. (credit a: modification of work by David Nance, USDA ARS; credit b, c: modification of work by Rosendahl; credit d: modification of work by Bill Tarpenning, USDA; credit e: modification of work by Scott Bauer, USDA ARS; credit f: modification of work by Keith Weller, USDA)

## Eudicots

Eudicots, or true dicots, are characterized by the presence of two cotyledons in the developing shoot. Veins form a network in leaves, and flower parts come in four, five, or many whorls. Vascular tissue forms a ring in the stem; in monocots, vascular tissue is scattered in the stem. Eudicots can be **herbaceous** (like grasses), or produce woody tissues. Most eudicots produce pollen that is trisulcate or triporate, with three furrows or pores.



The root system is usually anchored by one main root developed from the embryonic radicle. Eudicots comprise two-thirds of all flowering plants. The major differences between monocots and eudicots are summarized in [\[link\]](#). Many species exhibit characteristics that belong to either group; as such, the classification of a plant as a monocot or a eudicot is not always clearly evident.

Comparison of Structural Characteristics of Monocots and Eudicots		
Characteristic	Monocot	Eudicot
Cotyledon	One	Two
Veins in Leaves	Parallel	Network (branched)
Stem Vascular Tissue	Scattered	Arranged in ring pattern
Roots	Network of adventitious roots	Tap root with many lateral roots
Pollen	Monosulcate	Trisulcate
Flower Parts	Three or multiple of three	Four, five, multiple of four or five and whorls

Section Summary

Angiosperms are the dominant form of plant life in most terrestrial ecosystems, comprising about 90 percent of all plant species. Most crops and ornamental plants are angiosperms. Their success comes from two innovative structures that protect reproduction from variability in the environment: the flower and the fruit. Flowers were derived from modified leaves. The main parts of a flower are the sepals and petals, which protect the reproductive parts: the stamens and the carpels. The stamens produce the male gametes in pollen grains. The carpels contain the female gametes (the eggs inside the ovules), which are within the ovary of a carpel. The walls of the ovary thicken after fertilization, ripening into fruit that ensures dispersal by wind, water, or animals.

The angiosperm life cycle is dominated by the sporophyte stage. Double fertilization is an event unique to angiosperms. One sperm in the pollen fertilizes the egg, forming a diploid zygote, while the other combines with the two polar nuclei, forming a triploid cell that develops into a food storage tissue called the endosperm. Flowering plants are divided into two main groups, the monocots and eudicots, according to the number of cotyledons in the seedlings. Basal angiosperms belong to an older lineage than monocots and eudicots.

## Art Connections

### Exercise:

#### Problem:

[\[link\]](#) If a flower lacked a megasporangium, what type of gamete would not form? If the flower lacked a microsporangium, what type of gamete would not form?

---

#### Solution:

[\[link\]](#) Without a megasporangium, an egg would not form; without a microsporangium, pollen would not form.

## Review Questions

**Exercise:**

**Problem:**

Which of the following structures in a flower is not directly involved in reproduction?

- a. the style
- b. the stamen
- c. the sepal
- d. the anther

---

**Solution:**

C

**Exercise:**

**Problem:** Pollen grains develop in which structure?

- a. the anther
- b. the stigma
- c. the filament
- d. the carpel

---

**Solution:**

A

**Exercise:**

**Problem:**

In the course of double fertilization, one sperm cell fuses with the egg and the second one fuses with \_\_\_\_\_.

- a. the synergids
- b. the polar nuclei of the center cell

- c. the egg as well
- d. the antipodal cells

---

**Solution:**

B

**Exercise:**

**Problem:**

Corn develops from a seedling with a single cotyledon, displays parallel veins on its leaves, and produces monosulcate pollen. It is most likely:

- a. a gymnosperm
- b. a monocot
- c. a eudicot
- d. a basal angiosperm

---

**Solution:**

B

**Free Response**

**Exercise:**

**Problem:**

Some cycads are considered endangered species and their trade is severely restricted. Customs officials stop suspected smugglers who claim that the plants in their possession are palm trees, not cycads. How would a botanist distinguish between the two types of plants?

---

**Solution:**

The resemblance between cycads and palm trees is only superficial. Cycads are gymnosperms and do not bear flowers or fruit. Cycads produce cones: large, female cones that produce naked seeds, and smaller male cones on separate plants. Palms do not.

**Exercise:**

**Problem:**

What are the two structures that allow angiosperms to be the dominant form of plant life in most terrestrial ecosystems?

---

**Solution:**

Angiosperms are successful because of flowers and fruit. These structures protect reproduction from variability in the environment.

**Glossary**

anther

sac-like structure at the tip of the stamen in which pollen grains are produced

Anthophyta

phylum to which angiosperms belong

basal angiosperms

a group of plants that probably branched off before the separation of monocots and eudicots

calyx

whorl of sepals

carpel

single unit of the pistil

corolla

collection of petals

cotyledon

primitive leaf that develop in the zygote; monocots have one cotyledon, and dicots have two cotyledons

dicot

(also, eudicot) related group of angiosperms whose embryos possess two cotyledons

filament

thin stalk that links the anther to the base of the flower

gynoecium

(also, carpel) structure that constitute the female reproductive organ

herbaceous

grass-like plant noticeable by the absence of woody tissue

monocot

related group of angiosperms that produce embryos with one cotyledon and pollen with a single ridge

ovary

chamber that contains and protects the ovule or female megasporangium

perianth

part of the plant consisting of the calyx (sepals) and corolla (petals)

petal

modified leaf interior to the sepals; colorful petals attract animal pollinators

pistil

fused group of carpels

sepal

modified leaf that encloses the bud; outermost structure of a flower

stamen

structure that contains the male reproductive organs

stigma

uppermost structure of the carpel where pollen is deposited

style

long, thin structure that links the stigma to the ovary

## The Role of Seed Plants

By the end of this section, you will be able to:

- Explain how angiosperm diversity is due, in part, to multiple interactions with animals
- Describe ways in which pollination occurs
- Discuss the roles that plants play in ecosystems and how deforestation threatens plant biodiversity

Without seed plants, life as we know it would not be possible. Plants play a key role in the maintenance of terrestrial ecosystems through stabilization of soils, cycling of carbon, and climate moderation. Large tropical forests release oxygen and act as carbon dioxide sinks. Seed plants provide shelter to many life forms, as well as food for herbivores, thereby indirectly feeding carnivores. Plant secondary metabolites are used for medicinal purposes and industrial production.

## Animals and Plants: Herbivory

Coevolution of flowering plants and insects is a hypothesis that has received much attention and support, especially because both angiosperms and insects diversified at about the same time in the middle Mesozoic. Many authors have attributed the diversity of plants and insects to pollination and **herbivory**, or consumption of plants by insects and other animals. This is believed to have been as much a driving force as pollination. Coevolution of herbivores and plant defenses is observed in nature. Unlike animals, most plants cannot outrun predators or use mimicry to hide from hungry animals. A sort of arms race exists between plants and herbivores. To “combat” herbivores, some plant seeds—such as acorn and unripened persimmon—are high in alkaloids and therefore unsavory to some animals. Other plants are protected by bark, although some animals developed specialized mouth pieces to tear and chew vegetal material. Spines and thorns ([\[link\]](#)) deter most animals, except for mammals with thick fur, and some birds have specialized beaks to get past such defenses.





(a)

(b)

(a) Spines and (b) thorns are examples of plant defenses. (credit a: modification of work by Jon Sullivan; credit b: modification of work by I. Sáček, Sr.)

Herbivory has been used by seed plants for their own benefit in a display of mutualistic relationships. The dispersal of fruit by animals is the most striking example. The plant offers to the herbivore a nutritious source of food in return for spreading the plant's genetic material to a wider area.

An extreme example of collaboration between an animal and a plant is the case of acacia trees and ants. The trees support the insects with shelter and food. In return, ants discourage herbivores, both invertebrates and vertebrates, by stinging and attacking leaf-eating insects.

## **Animals and Plants: Pollination**

Grasses are a successful group of flowering plants that are wind pollinated. They produce large amounts of powdery pollen carried over large distances by the wind. The flowers are small and wisp-like. Large trees such as oaks, maples, and birches are also wind pollinated.

### **Note:**

[Link to Learning](#)



Explore this [website](#) for additional information on pollinators.

More than 80 percent of angiosperms depend on animals for **pollination**: the transfer of pollen from the anther to the stigma. Consequently, plants have developed many adaptations to attract pollinators. The specificity of specialized plant structures that target animals can be very surprising. It is possible, for example, to determine the type of pollinator favored by a plant just from the flower's characteristics. Many bird or insect-pollinated flowers secrete **nectar**, which is a sugary liquid. They also produce both fertile pollen, for reproduction, and sterile pollen rich in nutrients for birds and insects. Butterflies and bees can detect ultraviolet light. Flowers that attract these pollinators usually display a pattern of low ultraviolet reflectance that helps them quickly locate the flower's center and collect nectar while being dusted with pollen ([link](#)). Large, red flowers with little smell and a long funnel shape are preferred by hummingbirds, who have good color perception, a poor sense of smell, and need a strong perch. White flowers opened at night attract moths. Other animals—such as bats, lemurs, and lizards—can also act as pollinating agents. Any disruption to these interactions, such as the disappearance of bees as a consequence of colony collapse disorders, can lead to disaster for agricultural industries that depend heavily on pollinated crops.



As a bee collects nectar from a flower, it is dusted by pollen, which it then disperses to other flowers.  
(credit: John Severns)

**Note:**

**Scientific Method Connection**

**Testing Attraction of Flies by Rotting Flesh Smell**

Question: Will flowers that offer cues to bees attract carrion flies if sprayed with compounds that smell like rotten flesh?

Background: Visitation of flowers by pollinating flies is a function mostly of smell. Flies are attracted by rotting flesh and carrions. The putrid odor seems to be the major attractant. The polyamines putrescine and cadaverine, which are the products of protein breakdown after animal death, are the source of the pungent smell of decaying meat. Some plants strategically attract flies by synthesizing polyamines similar to those generated by decaying flesh and thereby attract carrion flies.

Flies seek out dead animals because they normally lay their eggs on them and their maggots feed on the decaying flesh. Interestingly, time of death can be determined by a forensic entomologist based on the stages and type of maggots recovered from cadavers.

Hypothesis: Because flies are drawn to other organisms based on smell and not sight, a flower that is normally attractive to bees because of its colors will attract flies if it is sprayed with polyamines similar to those generated by decaying flesh.

Test the hypothesis:

1. Select flowers usually pollinated by bees. White petunia may be good choice.
2. Divide the flowers into two groups, and while wearing eye protection and gloves, spray one group with a solution of either putrescine or cadaverine. (Putrescine dihydrochloride is typically available in 98 percent concentration; this can be diluted to approximately 50 percent for this experiment.)
3. Place the flowers in a location where flies are present, keeping the sprayed and unsprayed flowers separated.
4. Observe the movement of the flies for one hour. Record the number of visits to the flowers using a table similar to [\[link\]](#). Given the rapid movement of flies, it may be beneficial to use a video camera to record the fly–flower interaction. Replay the video in slow motion to obtain an accurate record of the number of fly visits to the flowers.
5. Repeat the experiment four more times with the same species of flower, but using different specimens.
6. Repeat the entire experiment with a different type of flower that is normally pollinated by bees.

### **Results of Number of Visits by Flies to Sprayed and Control/Unsprayed Flowers**

<b>Trial #</b>	<b>Sprayed Flowers</b>	<b>Unsprayed Flowers</b>
1		

### **Results of Number of Visits by Flies to Sprayed and Control/Unsprayed Flowers**

<b>Trial #</b>	<b>Sprayed Flowers</b>	<b>Unsprayed Flowers</b>
2		
3		
4		
5		

Analyze your data: Review the data you have recorded. Average the number of visits that flies made to sprayed flowers over the course of the five trials (on the first flower type) and compare and contrast them to the average number of visits that flies made to the unsprayed/control flowers. Can you draw any conclusions regarding the attraction of the flies to the sprayed flowers?

For the second flower type used, average the number of visits that flies made to sprayed flowers over the course of the five trials and compare and contrast them to the average number of visits that flies made to the unsprayed/control flowers. Can you draw any conclusions regarding the attraction of the flies to the sprayed flowers?

Compare and contrast the average number of visits that flies made to the two flower types. Can you draw any conclusions about whether the appearance of the flower had any impact on the attraction of flies? Did smell override any appearance differences, or were the flies attracted to one flower type more than another?

Form a conclusion: Do the results support the hypothesis? If not, how can this be explained?

## **The Importance of Seed Plants in Human Life**

Seed plants are the foundation of human diets across the world ([link](#)). Many societies eat almost exclusively vegetarian fare and depend solely on seed plants for their nutritional needs. A few **crops** (rice, wheat, and potatoes) dominate the agricultural landscape. Many crops were developed during the agricultural revolution, when human societies made the transition from nomadic hunter–gatherers to horticulture and agriculture. Cereals, rich in carbohydrates, provide the staple of many human diets. Beans and nuts supply proteins. Fats are derived from crushed seeds, as is the case for peanut and rapeseed (canola) oils, or fruits such as olives. Animal husbandry also consumes large amounts of crops.

Staple crops are not the only food derived from seed plants. Fruits and vegetables provide nutrients, vitamins, and fiber. Sugar, to sweeten dishes, is produced from the monocot sugarcane and the eudicot sugar beet. Drinks are made from infusions of tea leaves, chamomile flowers, crushed coffee beans, or powdered cocoa beans. Spices come from many different plant parts: saffron and cloves are stamens and buds, black pepper and vanilla are seeds, the bark of a bush in the *Laurales* family supplies cinnamon, and the herbs that flavor many dishes come from dried leaves and fruit, such as the pungent red chili pepper. The volatile oils of flowers and bark provide the scent of perfumes. Additionally, no discussion of seed plant contribution to human diet would be complete without the mention of alcohol. Fermentation of plant-derived sugars and starches is used to produce alcoholic beverages in all societies. In some cases, the beverages are derived from the fermentation of sugars from fruit, as with wines and, in other cases, from the fermentation of carbohydrates derived from seeds, as with beers.

Seed plants have many other uses, including providing wood as a source of timber for construction, fuel, and material to build furniture. Most paper is derived from the pulp of coniferous trees. Fibers of seed plants such as cotton, flax, and hemp are woven into cloth. Textile dyes, such as indigo, were mostly of plant origin until the advent of synthetic chemical dyes.

Lastly, it is more difficult to quantify the benefits of ornamental seed plants. These grace private and public spaces, adding beauty and serenity to human lives and inspiring painters and poets alike.



(a)



(b)



(c)



(d)

Humans rely on plants for a variety of reasons. (a) Cacao beans were introduced in Europe from the New World, where they were used by Mesoamerican civilizations. Combined with sugar, another plant product, chocolate is a popular food. (b) Flowers like the tulip are cultivated for their beauty. (c) Quinine, extracted from cinchona trees, is used to treat malaria, to reduce fever, and to alleviate pain. (d) This violin is made of wood. (credit a: modification of work by "Everjean"/Flickr; credit b:



modification of work by Rosendahl;  
credit c: modification of work by Franz  
Eugen Köhler)

The medicinal properties of plants have been known to human societies since ancient times. There are references to the use of plants' curative properties in Egyptian, Babylonian, and Chinese writings from 5,000 years ago. Many modern synthetic therapeutic drugs are derived or synthesized de novo from plant secondary metabolites. It is important to note that the same plant extract can be a therapeutic remedy at low concentrations, become an addictive drug at higher doses, and can potentially kill at high concentrations. [\[link\]](#) presents a few drugs, their plants of origin, and their medicinal applications.

Plant Origin of Medicinal Compounds and Medical Applications		
Plant	Compound	Application
Deadly nightshade ( <i>Atropa belladonna</i> )	Atropine	Dilate eye pupils for eye exams
Foxglove ( <i>Digitalis purpurea</i> )	Digitalis	Heart disease, stimulates heart beat
Yam ( <i>Dioscorea</i> spp.)	Steroids	Steroid hormones: contraceptive pill and cortisone



Plant Origin of Medicinal Compounds and Medical Applications		
Plant	Compound	Application
Ephedra ( <i>Ephedra</i> spp.)	Ephedrine	Decongestant and bronchiole dilator
Pacific yew ( <i>Taxus brevifolia</i> )	Taxol	Cancer chemotherapy; inhibits mitosis
Opium poppy ( <i>Papaver somniferum</i> )	Opioids	Analgesic (reduces pain without loss of consciousness) and narcotic (reduces pain with drowsiness and loss of consciousness) in higher doses
Quinine tree ( <i>Cinchona</i> spp.)	Quinine	Antipyretic (lowers body temperature) and antimalarial
Willow ( <i>Salix</i> spp.)	Salicylic acid (aspirin)	Analgesic and antipyretic

### Note:

#### Career Connection

#### Ethnobotanist

The relatively new field of ethnobotany studies the interaction between a particular culture and the plants native to the region. Seed plants have a large influence on day-to-day human life. Not only are plants the major source of food and medicine, they also influence many other aspects of society, from clothing to industry. The medicinal properties of plants were

recognized early on in human cultures. From the mid-1900s, synthetic chemicals began to supplant plant-based remedies.

Pharmacognosy is the branch of pharmacology that focuses on medicines derived from natural sources. With massive globalization and industrialization, there is a concern that much human knowledge of plants and their medicinal purposes will disappear with the cultures that fostered them. This is where ethnobotanists come in. To learn about and understand the use of plants in a particular culture, an ethnobotanist must bring in knowledge of plant life and an understanding and appreciation of diverse cultures and traditions. The Amazon forest is home to an incredible diversity of vegetation and is considered an untapped resource of medicinal plants; yet, both the ecosystem and its indigenous cultures are threatened with extinction.

To become an ethnobotanist, a person must acquire a broad knowledge of plant biology, ecology and sociology. Not only are the plant specimens studied and collected, but also the stories, recipes, and traditions that are linked to them. For ethnobotanists, plants are not viewed solely as biological organisms to be studied in a laboratory, but as an integral part of human culture. The convergence of molecular biology, anthropology, and ecology make the field of ethnobotany a truly multidisciplinary science.

## **Biodiversity of Plants**

Biodiversity ensures a resource for new food crops and medicines. Plant life balances ecosystems, protects watersheds, mitigates erosion, moderates climate and provides shelter for many animal species. Threats to plant diversity, however, come from many angles. The explosion of the human population, especially in tropical countries where birth rates are highest and economic development is in full swing, is leading to human encroachment into forested areas. To feed the larger population, humans need to obtain arable land, so there is massive clearing of trees. The need for more energy to power larger cities and economic growth therein leads to the construction of dams, the consequent flooding of ecosystems, and increased emissions of pollutants. Other threats to tropical forests come from poachers, who log trees for their precious wood. Ebony and Brazilian rosewood, both on the

endangered list, are examples of tree species driven almost to extinction by indiscriminate logging.

The number of plant species becoming extinct is increasing at an alarming rate. Because ecosystems are in a delicate balance, and seed plants maintain close symbiotic relationships with animals—whether predators or pollinators—the disappearance of a single plant can lead to the extinction of connected animal species. A real and pressing issue is that many plant species have not yet been catalogued, and so their place in the ecosystem is unknown. These unknown species are threatened by logging, habitat destruction, and loss of pollinators. They may become extinct before we have the chance to begin to understand the possible impacts from their disappearance. Efforts to preserve biodiversity take several lines of action, from preserving heirloom seeds to barcoding species. **Heirloom seeds** come from plants that were traditionally grown in human populations, as opposed to the seeds used for large-scale agricultural production.

**Barcoding** is a technique in which one or more short gene sequences, taken from a well-characterized portion of the genome, are used to identify a species through DNA analysis.

## Section Summary

Angiosperm diversity is due in part to multiple interactions with animals. Herbivory has favored the development of defense mechanisms in plants, and avoidance of those defense mechanism in animals. Pollination (the transfer of pollen to a carpel) is mainly carried out by wind and animals, and angiosperms have evolved numerous adaptations to capture the wind or attract specific classes of animals.

Plants play a key role in ecosystems. They are a source of food and medicinal compounds, and provide raw materials for many industries. Rapid deforestation and industrialization, however, threaten plant biodiversity. In turn, this threatens the ecosystem.

## Review Questions

### Exercise:

**Problem:**

Which of the following plant structures is not a defense against herbivory?

- a. thorns
- b. spines
- c. nectar
- d. alkaloids

---

**Solution:**

C

**Exercise:****Problem:**

White and sweet-smelling flowers with abundant nectar are probably pollinated by

- a. bees and butterflies
- b. flies
- c. birds
- d. wind

---

**Solution:**

A

**Exercise:****Problem:**

Abundant and powdery pollen produced by small, indistinct flowers is probably transported by:

- a. bees and butterflies

- b. flies
- c. birds
- d. wind

---

**Solution:**

D

**Exercise:**

**Problem:** Plants are a source of \_\_\_\_\_.

- a. food
- b. fuel
- c. medicine
- d. all of the above

---

**Solution:**

D

## Free Response

**Exercise:**

**Problem:**

Biosynthesis of nectar and nutrient-rich pollen is energetically very expensive for a plant. Yet, plants funnel large amounts of energy into animal pollination. What are the evolutionary advantages that offset the cost of attracting animal pollinators?

---

**Solution:**

Using animal pollinators promotes cross-pollination and increases genetic diversity. The odds that the pollen will reach another flower are

greatly increased compared with the randomness of wind pollination.

**Exercise:**

**Problem:**

What is biodiversity and why is it important to an ecosystem?

---

**Solution:**

Biodiversity is the variation in all forms of life. It can refer to variation within a species, within an ecosystem, or on an entire planet. It is important because it ensures a resource for new food crops and medicines. Plant life balances the ecosystems, protects watersheds, mitigates erosion, moderates climate, and provides shelter for many animal species.

## **Glossary**

barcoding

molecular biology technique in which one or more short gene sequences taken from a well-characterized portion of the genome is used to identify a species

crop

cultivated plant

heirloom seed

seed from a plant that was grown historically, but has not been used in modern agriculture on a large scale

herbivory

consumption of plants by insects and other animals

nectar

liquid rich in sugars produced by flowers to attract animal pollinators

pollination

transfer of pollen from the anther to the stigma

## Introduction

class="introduction"

A locust leaf consists of leaflets arrayed along a central midrib. Each leaflet is a complex photosynthetic machine, exquisitely adapted to capture sunlight and carbon dioxide. An intricate vascular system supplies the leaf with water and minerals, and exports the products of photosynthesis . (credit: modification of work by Todd Petit)





Plants are as essential to human existence as land, water, and air. Without plants, our day-to-day lives would be impossible because without oxygen from photosynthesis, aerobic life cannot be sustained. From providing food and shelter to serving as a source of medicines, oils, perfumes, and industrial products, plants provide humans with numerous valuable resources.

When you think of plants, most of the organisms that come to mind are vascular plants. These plants have tissues that conduct food and water, and they have seeds. Seed plants are divided into gymnosperms and angiosperms. Gymnosperms include the needle-leaved conifers—spruce, fir, and pine—as well as less familiar plants, such as ginkgos and cycads. Their seeds are not enclosed by a fleshy fruit. Angiosperms, also called flowering plants, constitute the majority of seed plants. They include broadleaved trees (such as maple, oak, and elm), vegetables (such as potatoes, lettuce, and carrots), grasses, and plants known for the beauty of their flowers (roses, irises, and daffodils, for example).

While individual plant species are unique, all share a common structure: a plant body consisting of stems, roots, and leaves. They all transport water, minerals, and sugars produced through photosynthesis through the plant body in a similar manner. All plant species also respond to environmental factors, such as light, gravity, competition, temperature, and predation.

## The Plant Body

By the end of this section, you will be able to:

- Describe the shoot organ system and the root organ system
- Distinguish between meristematic tissue and permanent tissue
- Identify and describe the three regions where plant growth occurs
- Summarize the roles of dermal tissue, vascular tissue, and ground tissue
- Compare simple plant tissue with complex plant tissue

Like animals, plants contain cells with organelles in which specific metabolic activities take place. Unlike animals, however, plants use energy from sunlight to form sugars during photosynthesis. In addition, plant cells have cell walls, plastids, and a large central vacuole: structures that are not found in animal cells. Each of these cellular structures plays a specific role in plant structure and function.

### Note:

Link to Learning

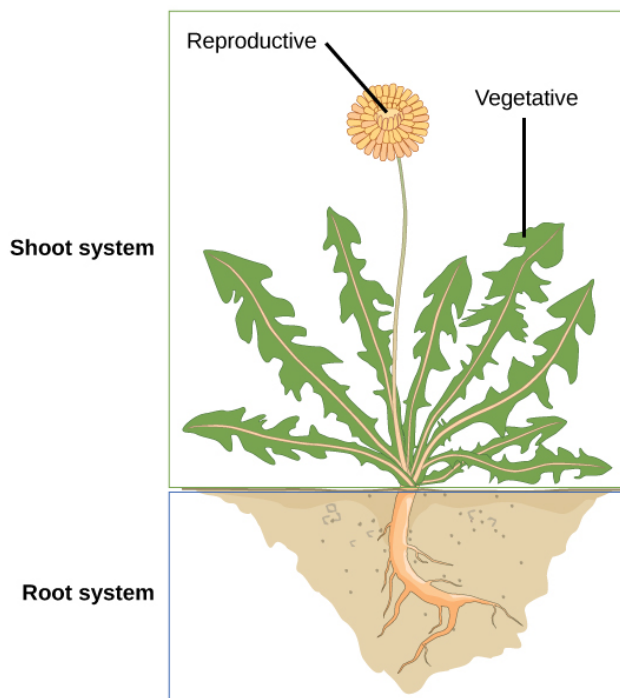


Watch [\*Botany Without Borders\*](#), a video produced by the Botanical Society of America about the importance of plants.

## Plant Organ Systems

In plants, just as in animals, similar cells working together form a tissue. When different types of tissues work together to perform a unique function, they form an organ; organs working together form organ systems. Vascular

plants have two distinct organ systems: a shoot system, and a root system. The **shoot system** consists of two portions: the vegetative (non-reproductive) parts of the plant, such as the leaves and the stems, and the reproductive parts of the plant, which include flowers and fruits. The shoot system generally grows above ground, where it absorbs the light needed for photosynthesis. The **root system**, which supports the plants and absorbs water and minerals, is usually underground. [\[link\]](#) shows the organ systems of a typical plant.



The shoot system of a plant consists of leaves, stems, flowers, and fruits. The root system anchors the plant while absorbing water and minerals from the soil.

## Plant Tissues

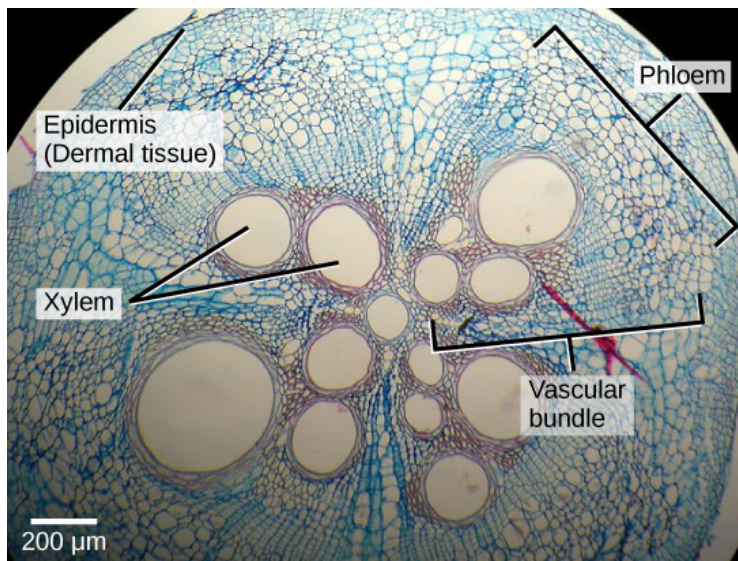
Plants are multicellular eukaryotes with tissue systems made of various cell types that carry out specific functions. Plant tissue systems fall into one of two general types: meristematic tissue, and permanent (or non-meristematic) tissue. Cells of the meristematic tissue are found in **meristems**, which are plant regions of continuous cell division and growth. **Meristematic tissue** cells are either undifferentiated or incompletely differentiated, and they continue to divide and contribute to the growth of the plant. In contrast, **permanent tissue** consists of plant cells that are no longer actively dividing.

Meristematic tissues consist of three types, based on their location in the plant. **Apical meristems** contain meristematic tissue located at the tips of stems and roots, which enable a plant to extend in length. **Lateral meristems** facilitate growth in thickness or girth in a maturing plant. **Intercalary meristems** occur only in monocots, at the bases of leaf blades and at nodes (the areas where leaves attach to a stem). This tissue enables the monocot leaf blade to increase in length from the leaf base; for example, it allows lawn grass leaves to elongate even after repeated mowing.

Meristems produce cells that quickly differentiate, or specialize, and become permanent tissue. Such cells take on specific roles and lose their ability to divide further. They differentiate into three main types: dermal, vascular, and ground tissue. **Dermal tissue** covers and protects the plant, and **vascular tissue** transports water, minerals, and sugars to different parts of the plant. **Ground tissue** serves as a site for photosynthesis, provides a supporting matrix for the vascular tissue, and helps to store water and sugars.

Secondary tissues are either simple (composed of similar cell types) or complex (composed of different cell types). Dermal tissue, for example, is a simple tissue that covers the outer surface of the plant and controls gas exchange. Vascular tissue is an example of a complex tissue, and is made of two specialized conducting tissues: xylem and phloem. Xylem tissue transports water and nutrients from the roots to different parts of the plant, and includes three different cell types: vessel elements and tracheids (both of which conduct water), and xylem parenchyma. Phloem tissue, which transports organic compounds from the site of photosynthesis to other parts

of the plant, consists of four different cell types: sieve cells (which conduct photosynthates), companion cells, phloem parenchyma, and phloem fibers. Unlike xylem conducting cells, phloem conducting cells are alive at maturity. The xylem and phloem always lie adjacent to each other ([\[link\]](#)). In stems, the xylem and the phloem form a structure called a **vascular bundle**; in roots, this is termed the **vascular stele** or **vascular cylinder**.



This light micrograph shows a cross section of a squash (*Cucurbita maxima*) stem. Each teardrop-shaped vascular bundle consists of large xylem vessels toward the inside and smaller phloem cells toward the outside. Xylem cells, which transport water and nutrients from the roots to the rest of the plant, are dead at functional maturity. Phloem cells, which transport sugars and other organic compounds from photosynthetic tissue to the rest of the plant, are living. The vascular bundles are encased in ground tissue and surrounded by dermal tissue.  
(credit: modification of work by "

(biophotos)"/Flickr; scale-bar data from  
Matt Russell)

## Section Summary

A vascular plant consists of two organ systems: the shoot system and the root system. The shoot system includes the aboveground vegetative portions (stems and leaves) and reproductive parts (flowers and fruits). The root system supports the plant and is usually underground. A plant is composed of two main types of tissue: meristematic tissue and permanent tissue. Meristematic tissue consists of actively dividing cells found in root and shoot tips. As growth occurs, meristematic tissue differentiates into permanent tissue, which is categorized as either simple or complex. Simple tissues are made up of similar cell types; examples include dermal tissue and ground tissue. Dermal tissue provides the outer covering of the plant. Ground tissue is responsible for photosynthesis; it also supports vascular tissue and may store water and sugars. Complex tissues are made up of different cell types. Vascular tissue, for example, is made up of xylem and phloem cells.

## Review Questions

### Exercise:

#### Problem:

Plant regions of continuous growth are made up of \_\_\_\_\_.

- a. dermal tissue
- b. vascular tissue
- c. meristematic tissue
- d. permanent tissue

---

#### Solution:

C

**Exercise:**

**Problem:** Which of the following is the major site of photosynthesis?

- a. apical meristem
  - b. ground tissue
  - c. xylem cells
  - d. phloem cells
- 

**Solution:**

B

**Free Response**

**Exercise:**

**Problem:**

What type of meristem is found only in monocots, such as lawn grasses? Explain how this type of meristematic tissue is beneficial in lawn grasses that are mowed each week.

---

**Solution:**

Lawn grasses and other monocots have an intercalary meristem, which is a region of meristematic tissue at the base of the leaf blade. This is beneficial to the plant because it can continue to grow even when the tip of the plant is removed by grazing or mowing.

**Exercise:**

**Problem:**

Which plant part is responsible for transporting water, minerals, and sugars to different parts of the plant? Name the two types of tissue that make up this overall tissue, and explain the role of each.

---

**Solution:**

Vascular tissue transports water, minerals, and sugars throughout the plant. Vascular tissue is made up of xylem tissue and phloem tissue. Xylem tissue transports water and nutrients from the roots upward. Phloem tissue carries sugars from the sites of photosynthesis to the rest of the plant.

**Glossary**

apical meristem

meristematic tissue located at the tips of stems and roots; enables a plant to extend in length

dermal tissue

protective plant tissue covering the outermost part of the plant; controls gas exchange

ground tissue

plant tissue involved in photosynthesis; provides support, and stores water and sugars

intercalary meristem

meristematic tissue located at nodes and the bases of leaf blades; found only in monocots

lateral meristem

meristematic tissue that enables a plant to increase in thickness or girth

meristematic tissue



tissue containing cells that constantly divide; contributes to plant growth

meristem

plant region of continuous growth

permanent tissue

plant tissue composed of cells that are no longer actively dividing

root system

belowground portion of the plant that supports the plant and absorbs water and minerals

shoot system

aboveground portion of the plant; consists of non-reproductive plant parts, such as leaves and stems, and reproductive parts, such as flowers and fruits

vascular bundle

strands of stem tissue made up of xylem and phloem

vascular stele

strands of root tissue made up of xylem and phloem

vascular tissue

tissue made up of xylem and phloem that transports food and water throughout the plant

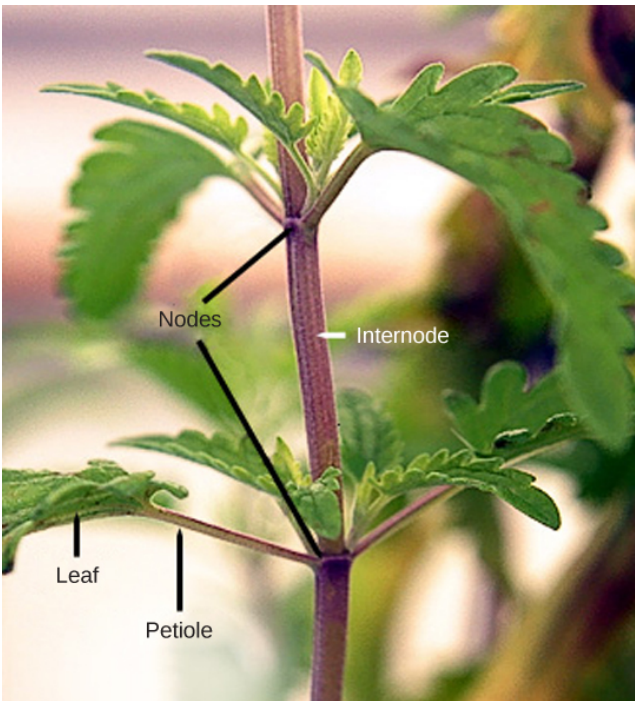
## Stems

By the end of this section, you will be able to:

- Describe the main function and basic structure of stems
- Compare and contrast the roles of dermal tissue, vascular tissue, and ground tissue
- Distinguish between primary growth and secondary growth in stems
- Summarize the origin of annual rings
- List and describe examples of modified stems

Stems are a part of the shoot system of a plant. They may range in length from a few millimeters to hundreds of meters, and also vary in diameter, depending on the plant type. Stems are usually above ground, although the stems of some plants, such as the potato, also grow underground. Stems may be herbaceous (soft) or woody in nature. Their main function is to provide support to the plant, holding leaves, flowers and buds; in some cases, stems also store food for the plant. A stem may be unbranched, like that of a palm tree, or it may be highly branched, like that of a magnolia tree. The stem of the plant connects the roots to the leaves, helping to transport absorbed water and minerals to different parts of the plant. It also helps to transport the products of photosynthesis, namely sugars, from the leaves to the rest of the plant.

Plant stems, whether above or below ground, are characterized by the presence of nodes and internodes ([\[link\]](#)). **Nodes** are points of attachment for leaves, aerial roots, and flowers. The stem region between two nodes is called an **internode**. The stalk that extends from the stem to the base of the leaf is the petiole. An **axillary bud** is usually found in the axil—the area between the base of a leaf and the stem—where it can give rise to a branch or a flower. The apex (tip) of the shoot contains the apical meristem within the **apical bud**.



Leaves are attached to the plant stem at areas called nodes. An internode is the stem region between two nodes. The petiole is the stalk connecting the leaf to the stem. The leaves just above the nodes arose from axillary buds.

## Stem Anatomy

The stem and other plant organs arise from the ground tissue, and are primarily made up of simple tissues formed from three types of cells: parenchyma, collenchyma, and sclerenchyma cells.

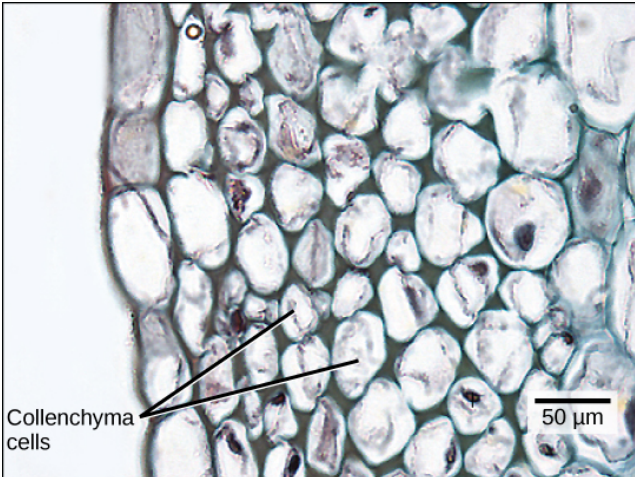
**Parenchyma cells** are the most common plant cells ([\[link\]](#)). They are found in the stem, the root, the inside of the leaf, and the pulp of the fruit. Parenchyma cells are responsible for metabolic functions, such as

photosynthesis, and they help repair and heal wounds. Some parenchyma cells also store starch.



The stem of common St John's Wort (*Hypericum perforatum*) is shown in cross section in this light micrograph. The central pith (greenish-blue, in the center) and peripheral cortex (narrow zone 3–5 cells thick just inside the epidermis) are composed of parenchyma cells. Vascular tissue composed of xylem (red) and phloem tissue (green, between the xylem and cortex) surrounds the pith. (credit: Rolf-Dieter Mueller)

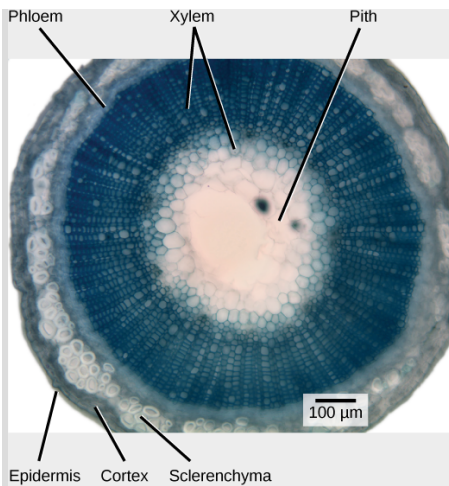
**Collenchyma cells** are elongated cells with unevenly thickened walls ([link](#)). They provide structural support, mainly to the stem and leaves. These cells are alive at maturity and are usually found below the epidermis. The “strings” of a celery stalk are an example of collenchyma cells.



Collenchyma cell walls are uneven in thickness, as seen in this light micrograph. They provide support to plant structures. (credit: modification of work by Carl Szczerski; scale-bar data from Matt Russell)

**Sclerenchyma cells** also provide support to the plant, but unlike collenchyma cells, many of them are dead at maturity. There are two types of sclerenchyma cells: fibers and sclereids. Both types have secondary cell walls that are thickened with deposits of lignin, an organic compound that is a key component of wood. Fibers are long, slender cells; sclereids are smaller-sized. Sclereids give pears their gritty texture. Humans use sclerenchyma fibers to make linen and rope ([link](#)).

**Note:**  
Art Connection



(a)



(b)



(c)

The central pith and outer cortex of the (a) flax stem are made up of parenchyma cells. Inside the cortex is a layer of sclerenchyma cells, which make up the fibers in flax rope and clothing.

Humans have grown and harvested flax for thousands of years. In (b) this drawing, fourteenth-century women prepare linen. The (c) flax plant is grown and harvested for its fibers, which are used to weave linen, and for its seeds, which are the source of linseed oil. (credit a: modification of work by Emmanuel Boutet based on original work by Ryan R. MacKenzie; credit c: modification of work by Brian Dearth; scale-bar data from Matt Russell)

Which layers of the stem are made of parenchyma cells?

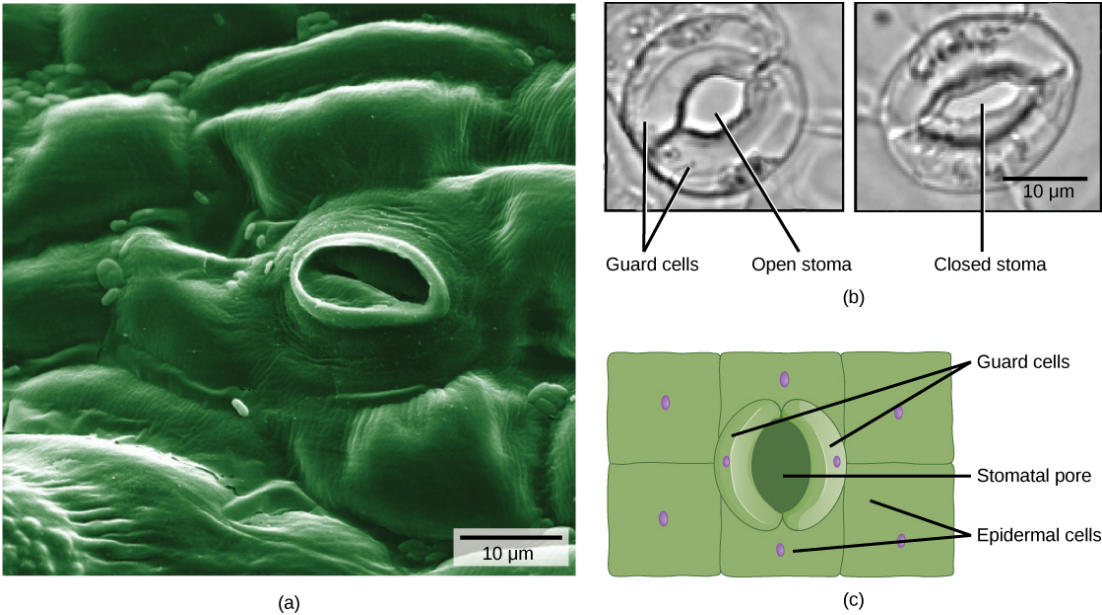


- a. cortex and pith
- b. phloem
- c. sclerenchyma
- d. xylem

Like the rest of the plant, the stem has three tissue systems: dermal, vascular, and ground tissue. Each is distinguished by characteristic cell types that perform specific tasks necessary for the plant's growth and survival.

## Dermal Tissue

The dermal tissue of the stem consists primarily of **epidermis**, a single layer of cells covering and protecting the underlying tissue. Woody plants have a tough, waterproof outer layer of cork cells commonly known as **bark**, which further protects the plant from damage. Epidermal cells are the most numerous and least differentiated of the cells in the epidermis. The epidermis of a leaf also contains openings known as stomata, through which the exchange of gases takes place ([\[link\]](#)). Two cells, known as **guard cells**, surround each leaf stoma, controlling its opening and closing and thus regulating the uptake of carbon dioxide and the release of oxygen and water vapor. **Trichomes** are hair-like structures on the epidermal surface. They help to reduce **transpiration** (the loss of water by aboveground plant parts), increase solar reflectance, and store compounds that defend the leaves against predation by herbivores.



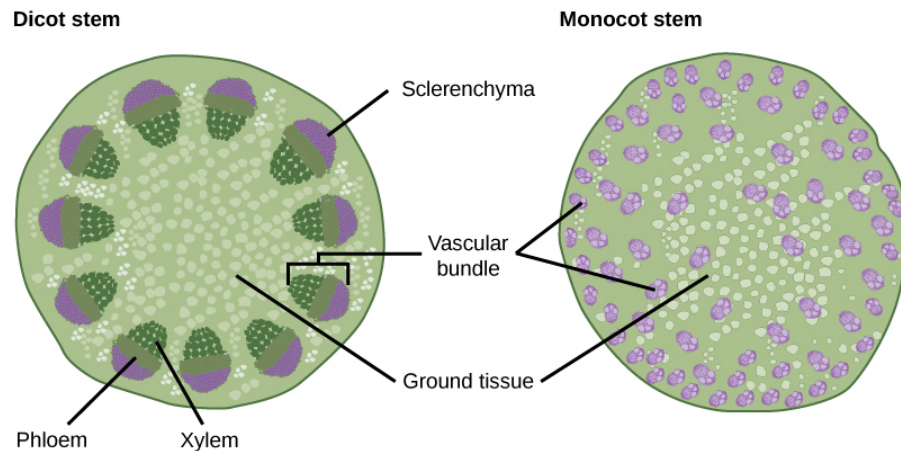
Openings called stomata (singular: stoma) allow a plant to take up carbon dioxide and release oxygen and water vapor. The (a) colorized scanning-electron micrograph shows a closed stoma of a dicot. Each stoma is flanked by two guard cells that regulate its (b) opening and closing. The (c) guard cells sit within the layer of epidermal cells (credit a: modification of work by Louisa Howard, Rippel Electron Microscope Facility, Dartmouth College; credit b: modification of work by June Kwak, University of Maryland; scale-bar data from Matt Russell)

## Vascular Tissue

The xylem and phloem that make up the vascular tissue of the stem are arranged in distinct strands called vascular bundles, which run up and down the length of the stem. When the stem is viewed in cross section, the vascular bundles of dicot stems are arranged in a ring. In plants with stems that live for more than one year, the individual bundles grow together and



produce the characteristic growth rings. In monocot stems, the vascular bundles are randomly scattered throughout the ground tissue ([link](#)).



In (a) dicot stems, vascular bundles are arranged around the periphery of the ground tissue. The xylem tissue is located toward the interior of the vascular bundle, and phloem is located toward the exterior. Sclerenchyma fibers cap the vascular bundles. In (b) monocot stems, vascular bundles composed of xylem and phloem tissues are scattered throughout the ground tissue.

Xylem tissue has three types of cells: xylem parenchyma, tracheids, and vessel elements. The latter two types conduct water and are dead at maturity. **Tracheids** are xylem cells with thick secondary cell walls that are lignified. Water moves from one tracheid to another through regions on the side walls known as pits, where secondary walls are absent. **Vessel elements** are xylem cells with thinner walls; they are shorter than tracheids. Each vessel element is connected to the next by means of a perforation plate at the end walls of the element. Water moves through the perforation plates to travel up the plant.

Phloem tissue is composed of sieve-tube cells, companion cells, phloem parenchyma, and phloem fibers. A series of **sieve-tube cells** (also called sieve-tube elements) are arranged end to end to make up a long sieve tube, which transports organic substances such as sugars and amino acids. The sugars flow from one sieve-tube cell to the next through perforated sieve plates, which are found at the end junctions between two cells. Although still alive at maturity, the nucleus and other cell components of the sieve-tube cells have disintegrated. **Companion cells** are found alongside the sieve-tube cells, providing them with metabolic support. The companion cells contain more ribosomes and mitochondria than the sieve-tube cells, which lack some cellular organelles.

## Ground Tissue

Ground tissue is mostly made up of parenchyma cells, but may also contain collenchyma and sclerenchyma cells that help support the stem. The ground tissue towards the interior of the vascular tissue in a stem or root is known as **pith**, while the layer of tissue between the vascular tissue and the epidermis is known as the **cortex**.

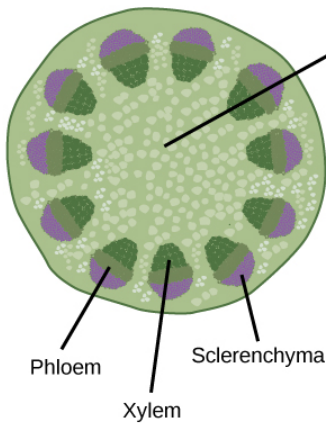
## Growth in Stems

Growth in plants occurs as the stems and roots lengthen. Some plants, especially those that are woody, also increase in thickness during their life span. The increase in length of the shoot and the root is referred to as **primary growth**, and is the result of cell division in the shoot apical meristem. **Secondary growth** is characterized by an increase in thickness or girth of the plant, and is caused by cell division in the lateral meristem.

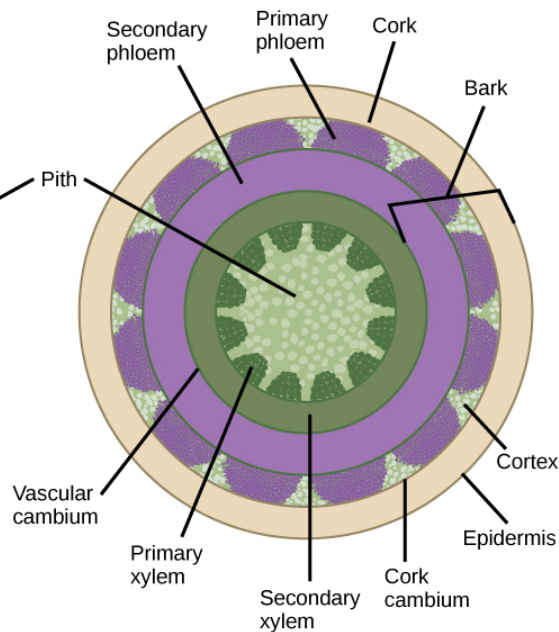
[\[link\]](#) shows the areas of primary and secondary growth in a plant.

Herbaceous plants mostly undergo primary growth, with hardly any secondary growth or increase in thickness. Secondary growth or “wood” is noticeable in woody plants; it occurs in some dicots, but occurs very rarely in monocots.

### Primary growth



### Secondary growth



In woody plants, primary growth is followed by secondary growth, which allows the plant stem to increase in thickness or girth. Secondary vascular tissue is added as the plant grows, as well as a cork layer. The bark of a tree extends from the vascular cambium to the epidermis.

Some plant parts, such as stems and roots, continue to grow throughout a plant's life: a phenomenon called indeterminate growth. Other plant parts, such as leaves and flowers, exhibit determinate growth, which ceases when a plant part reaches a particular size.

## Primary Growth

Most primary growth occurs at the apices, or tips, of stems and roots. Primary growth is a result of rapidly dividing cells in the apical meristems at the shoot tip and root tip. Subsequent cell elongation also contributes to

primary growth. The growth of shoots and roots during primary growth enables plants to continuously seek water (roots) or sunlight (shoots).

The influence of the apical bud on overall plant growth is known as apical dominance, which diminishes the growth of axillary buds that form along the sides of branches and stems. Most coniferous trees exhibit strong apical dominance, thus producing the typical conical Christmas tree shape. If the apical bud is removed, then the axillary buds will start forming lateral branches. Gardeners make use of this fact when they prune plants by cutting off the tops of branches, thus encouraging the axillary buds to grow out, giving the plant a bushy shape.

**Note:**

Link to Learning



Watch this [BBC Nature video](#) showing how time-lapse photography captures plant growth at high speed.

## Secondary Growth

The increase in stem thickness that results from secondary growth is due to the activity of the lateral meristems, which are lacking in herbaceous plants. Lateral meristems include the vascular cambium and, in woody plants, the cork cambium (see [link](#)). The vascular cambium is located just outside the primary xylem and to the interior of the primary phloem. The cells of the vascular cambium divide and form secondary xylem (tracheids and vessel elements) to the inside, and secondary phloem (sieve elements and companion cells) to the outside. The thickening of the stem that occurs in

secondary growth is due to the formation of secondary phloem and secondary xylem by the vascular cambium, plus the action of cork cambium, which forms the tough outermost layer of the stem. The cells of the secondary xylem contain lignin, which provides hardness and strength.

In woody plants, cork cambium is the outermost lateral meristem. It produces cork cells (bark) containing a waxy substance known as suberin that can repel water. The bark protects the plant against physical damage and helps reduce water loss. The cork cambium also produces a layer of cells known as phelloderm, which grows inward from the cambium. The cork cambium, cork cells, and phelloderm are collectively termed the **periderm**. The periderm substitutes for the epidermis in mature plants. In some plants, the periderm has many openings, known as **lenticels**, which allow the interior cells to exchange gases with the outside atmosphere ([link](#)). This supplies oxygen to the living and metabolically active cells of the cortex, xylem and phloem.

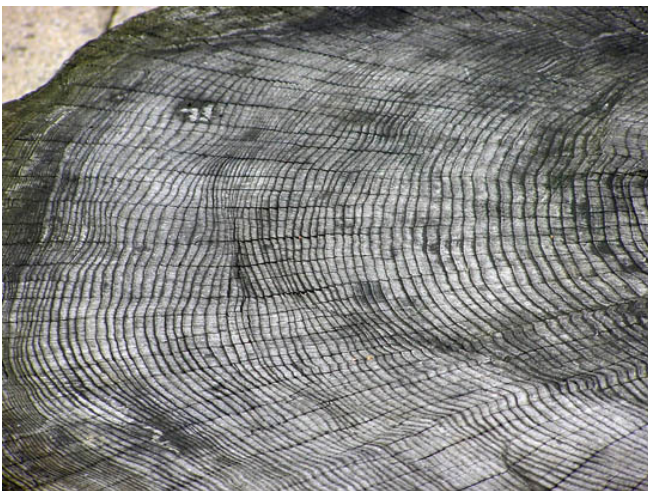


Lenticels on the bark of this cherry tree enable the woody stem to exchange gases with

the surrounding atmosphere.  
(credit: Roger Griffith)

## Annual Rings

The activity of the vascular cambium gives rise to annual growth rings. During the spring growing season, cells of the secondary xylem have a large internal diameter and their primary cell walls are not extensively thickened. This is known as early wood, or spring wood. During the fall season, the secondary xylem develops thickened cell walls, forming late wood, or autumn wood, which is denser than early wood. This alternation of early and late wood is due largely to a seasonal decrease in the number of vessel elements and a seasonal increase in the number of tracheids. It results in the formation of an annual ring, which can be seen as a circular ring in the cross section of the stem ([link](#)). An examination of the number of annual rings and their nature (such as their size and cell wall thickness) can reveal the age of the tree and the prevailing climatic conditions during each season.



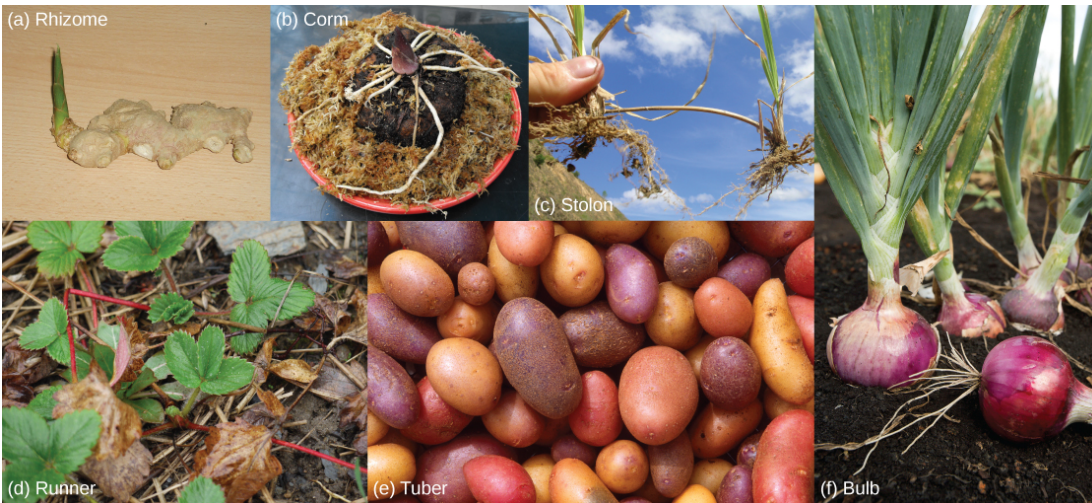
The rate of wood growth increases



in summer and decreases in winter, producing a characteristic ring for each year of growth. Seasonal changes in weather patterns can also affect the growth rate—note how the rings vary in thickness.  
(credit: Adrian Pingstone)

## Stem Modifications

Some plant species have modified stems that are especially suited to a particular habitat and environment ([link](#)). A **rhizome** is a modified stem that grows horizontally underground and has nodes and internodes. Vertical shoots may arise from the buds on the rhizome of some plants, such as ginger and ferns. **Corms** are similar to rhizomes, except they are more rounded and fleshy (such as in gladiolus). Corms contain stored food that enables some plants to survive the winter. **Stolons** are stems that run almost parallel to the ground, or just below the surface, and can give rise to new plants at the nodes. **Runners** are a type of stolon that runs above the ground and produces new clone plants at nodes at varying intervals: strawberries are an example. **Tubers** are modified stems that may store starch, as seen in the potato (*Solanum* sp.). Tubers arise as swollen ends of stolons, and contain many adventitious or unusual buds (familiar to us as the “eyes” on potatoes). A **bulb**, which functions as an underground storage unit, is a modification of a stem that has the appearance of enlarged fleshy leaves emerging from the stem or surrounding the base of the stem, as seen in the iris.



Stem modifications enable plants to thrive in a variety of environments. Shown are (a) ginger (*Zingiber officinale*) rhizomes, (b) a carrion flower (*Amorphophallus titanum*) corm (c) Rhodes grass (*Chloris gayana*) stolons, (d) strawberry (*Fragaria ananassa*) runners, (e) potato (*Solanum tuberosum*) tubers, and (f) red onion (*Allium*) bulbs. (credit a: modification of work by Maja Dumat; credit c: modification of work by Harry Rose; credit d: modification of work by Rebecca Siegel; credit e: modification of work by Scott Bauer, USDA ARS; credit f: modification of work by Stephen Ausmus, USDA ARS)

### Note:

Link to Learning





Watch botanist Wendy Hodgson, of Desert Botanical Garden in Phoenix, Arizona, explain how agave plants were cultivated for food hundreds of years ago in the Arizona desert in this [video](#): *Finding the Roots of an Ancient Crop*.

Some aerial modifications of stems are tendrils and thorns ([link](#)). **Tendrils** are slender, twining strands that enable a plant (like a vine or pumpkin) to seek support by climbing on other surfaces. **Thorns** are modified branches appearing as sharp outgrowths that protect the plant; common examples include roses, Osage orange and devil's walking stick.



(a)

(b)

Found in southeastern United States, (a) buckwheat vine (*Brunnichia ovata*) is a weedy plant that climbs with the aid of tendrils. This one is shown climbing up a wooden stake. (b) Thorns are modified branches. (credit a: modification of work by Christopher Meloche, USDA ARS; credit b: modification of work by “macrophile”/Flickr)

## Section Summary

The stem of a plant bears the leaves, flowers, and fruits. Stems are characterized by the presence of nodes (the points of attachment for leaves or branches) and internodes (regions between nodes).

Plant organs are made up of simple and complex tissues. The stem has three tissue systems: dermal, vascular, and ground tissue. Dermal tissue is the outer covering of the plant. It contains epidermal cells, stomata, guard cells, and trichomes. Vascular tissue is made up of xylem and phloem tissues and conducts water, minerals, and photosynthetic products. Ground tissue is responsible for photosynthesis and support and is composed of parenchyma, collenchyma, and sclerenchyma cells.

Primary growth occurs at the tips of roots and shoots, causing an increase in length. Woody plants may also exhibit secondary growth, or increase in thickness. In woody plants, especially trees, annual rings may form as growth slows at the end of each season. Some plant species have modified stems that help to store food, propagate new plants, or discourage predators. Rhizomes, corms, stolons, runners, tubers, bulbs, tendrils, and thorns are examples of modified stems.

## Art Connections

### Exercise:

#### Problem:

[\[link\]](#) Which layers of the stem are made of parenchyma cells?

- A. cortex and pith
- B. epidermis
- C. sclerenchyma
- D. epidermis and cortex.

---

#### Solution:

[\[link\]](#) A and B. The cortex, pith, and epidermis are made of parenchyma cells.

## Review Questions

### Exercise:

#### Problem:

Stem regions at which leaves are attached are called \_\_\_\_\_.

- a. trichomes
- b. lenticels
- c. nodes
- d. internodes

---

#### Solution:

C

### Exercise:

#### Problem:

Which of the following cell types forms most of the inside of a plant?

- a. meristem cells
- b. collenchyma cells
- c. sclerenchyma cells
- d. parenchyma cells

---

#### Solution:

D

### Exercise:

**Problem:**

Tracheids, vessel elements, sieve-tube cells, and companion cells are components of \_\_\_\_\_.

- a. vascular tissue
- b. meristematic tissue
- c. ground tissue
- d. dermal tissue

---

**Solution:**

A

**Exercise:****Problem:**

The primary growth of a plant is due to the action of the \_\_\_\_\_.

- a. lateral meristem
- b. vascular cambium
- c. apical meristem
- d. cork cambium

---

**Solution:**

C

**Exercise:**

**Problem:** Which of the following is an example of secondary growth?

- a. increase in length
- b. increase in thickness or girth
- c. increase in root hairs
- d. increase in leaf number

---

**Solution:**

B

**Exercise:**

**Problem:** Secondary growth in stems is usually seen in \_\_\_\_\_.

- a. monocots
- b. dicots
- c. both monocots and dicots
- d. neither monocots nor dicots

---

**Solution:**

B

**Free Response****Exercise:****Problem:**

Describe the roles played by stomata and guard cells. What would happen to a plant if these cells did not function correctly?

---

**Solution:**

Stomata allow gases to enter and exit the plant. Guard cells regulate the opening and closing of stomata. If these cells did not function correctly, a plant could not get the carbon dioxide needed for photosynthesis, nor could it release the oxygen produced by photosynthesis.

**Exercise:**

**Problem:**

Compare the structure and function of xylem to that of phloem.

---

**Solution:**

Xylem is made up of tracheids and vessel elements, which are cells that transport water and dissolved minerals and that are dead at maturity. Phloem is made up of sieve-tube cells and companion cells, which transport carbohydrates and are alive at maturity.

**Exercise:**

**Problem:** Explain the role of the cork cambium in woody plants.

---

**Solution:**

In woody plants, the cork cambium is the outermost lateral meristem; it produces new cells towards the interior, which enables the plant to increase in girth. The cork cambium also produces cork cells towards the exterior, which protect the plant from physical damage while reducing water loss.

**Exercise:**

**Problem:** What is the function of lenticels?

---

**Solution:**

In woody stems, lenticels allow internal cells to exchange gases with the outside atmosphere.

**Exercise:****Problem:**

Besides the age of a tree, what additional information can annual rings reveal?

---

**Solution:**

Annual rings can also indicate the climate conditions that prevailed during each growing season.

**Exercise:****Problem:**

Give two examples of modified stems and explain how each example benefits the plant.

---

**Solution:**

Answers will vary. Rhizomes, stolons, and runners can give rise to new plants. Corms, tubers, and bulbs can also produce new plants and can store food. Tendrils help a plant to climb, while thorns discourage herbivores.

**Glossary**

apical bud

bud formed at the tip of the shoot

axillary bud

bud located in the axil: the stem area where the petiole connects to the stem

bark

tough, waterproof, outer epidermal layer of cork cells

bulb

modified underground stem that consists of a large bud surrounded by numerous leaf scales

collenchyma cell

elongated plant cell with unevenly thickened walls; provides structural support to the stem and leaves

companion cell

phloem cell that is connected to sieve-tube cells; has large amounts of ribosomes and mitochondrion

corm

rounded, fleshy underground stem that contains stored food

cortex

ground tissue found between the vascular tissue and the epidermis in a stem or root

epidermis

single layer of cells found in plant dermal tissue; covers and protects underlying tissue

guard cells

paired cells on either side of a stoma that control stomatal opening and thereby regulate the movement of gases and water vapor

internode

region between nodes on the stem

lenticel

opening on the surface of mature woody stems that facilitates gas exchange

node

point along the stem at which leaves, flowers, or aerial roots originate

parenchyma cell

most common type of plant cell; found in the stem, root, leaf, and in fruit pulp; site of photosynthesis and starch storage

periderm

outermost covering of woody stems; consists of the cork cambium, cork cells, and the phelloderm

pith



ground tissue found towards the interior of the vascular tissue in a stem or root

primary growth

growth resulting in an increase in length of the stem and the root; caused by cell division in the shoot or root apical meristem

rhizome

modified underground stem that grows horizontally to the soil surface and has nodes and internodes

runner

stolon that runs above the ground and produces new clone plants at nodes

sclerenchyma cell

plant cell that has thick secondary walls and provides structural support; usually dead at maturity

secondary growth

growth resulting in an increase in thickness or girth; caused by the lateral meristem and cork cambium

sieve-tube cell

phloem cell arranged end to end to form a sieve tube that transports organic substances such as sugars and amino acids

stolon

modified stem that runs parallel to the ground and can give rise to new plants at the nodes

tendrils

modified stem consisting of slender, twining strands used for support or climbing

thorn

modified stem branch appearing as a sharp outgrowth that protects the plant

tracheid

xylem cell with thick secondary walls that helps transport water

trichome

hair-like structure on the epidermal surface

tuber

modified underground stem adapted for starch storage; has many adventitious buds

vessel element

xylem cell that is shorter than a tracheid and has thinner walls

## Roots

By the end of this section, you will be able to:

- Identify the two types of root systems
- Describe the three zones of the root tip and summarize the role of each zone in root growth
- Describe the structure of the root
- List and describe examples of modified roots

The roots of seed plants have three major functions: anchoring the plant to the soil, absorbing water and minerals and transporting them upwards, and storing the products of photosynthesis. Some roots are modified to absorb moisture and exchange gases. Most roots are underground. Some plants, however, also have **adventitious roots**, which emerge above the ground from the shoot.

## Types of Root Systems

Root systems are mainly of two types ([\[link\]](#)). Dicots have a tap root system, while monocots have a fibrous root system. A **tap root system** has a main root that grows down vertically, and from which many smaller lateral roots arise. Dandelions are a good example; their tap roots usually break off when trying to pull these weeds, and they can regrow another shoot from the remaining root). A tap root system penetrates deep into the soil. In contrast, a **fibrous root system** is located closer to the soil surface, and forms a dense network of roots that also helps prevent soil erosion (lawn grasses are a good example, as are wheat, rice, and corn). Some plants have a combination of tap roots and fibrous roots. Plants that grow in dry areas often have deep root systems, whereas plants growing in areas with abundant water are likely to have shallower root systems.

(a) Taproot system



(b) Fibrous root system

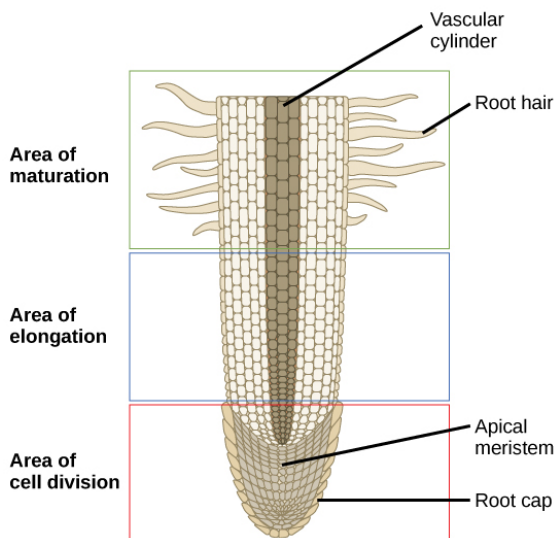


(a) Tap root systems have a main root that grows down, while (b) fibrous root systems consist of many small roots. (credit b: modification of work by “Austin Squarepants”/Flickr)

## Root Growth and Anatomy

Root growth begins with seed germination. When the plant embryo emerges from the seed, the radicle of the embryo forms the root system. The tip of the root is protected by the **root cap**, a structure exclusive to roots and unlike any other plant structure. The root cap is continuously replaced because it gets damaged easily as the root pushes through soil. The root tip can be divided into three zones: a zone of cell division, a zone of elongation, and a zone of maturation and differentiation ([\[link\]](#)). The zone of cell division is closest to the root tip; it is made up of the actively dividing cells of the root meristem. The zone of elongation is where the newly formed cells increase in length, thereby lengthening the root. Beginning at the first root hair is the zone of cell maturation where the root

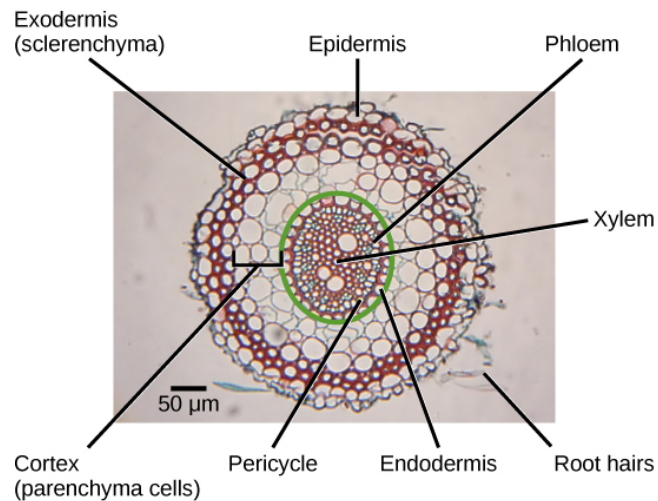
cells begin to differentiate into special cell types. All three zones are in the first centimeter or so of the root tip.



A longitudinal view of the root reveals the zones of cell division, elongation, and maturation. Cell division occurs in the apical meristem.

The root has an outer layer of cells called the epidermis, which surrounds areas of ground tissue and vascular tissue. The epidermis provides protection and helps in absorption. **Root hairs**, which are extensions of root epidermal cells, increase the surface area of the root, greatly contributing to the absorption of water and minerals.

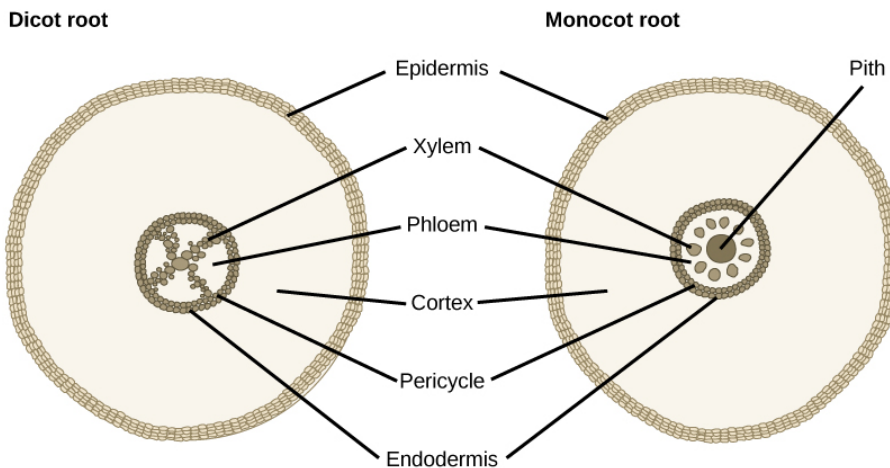
Inside the root, the ground tissue forms two regions: the cortex and the pith ([link](#)). Compared to stems, roots have lots of cortex and little pith. Both regions include cells that store photosynthetic products. The cortex is between the epidermis and the vascular tissue, whereas the pith lies between the vascular tissue and the center of the root.



Staining reveals different cell types in this light micrograph of a wheat (*Triticum*) root cross section. Sclerenchyma cells of the exodermis and xylem cells stain red, and phloem cells stain blue. Other cell types stain black. The stele, or vascular tissue, is the area inside endodermis (indicated by a green ring). Root hairs are visible outside the epidermis. (credit: scale-bar data from Matt Russell)

The vascular tissue in the root is arranged in the inner portion of the root, which is called the **stele** ([link](#)). A layer of cells known as the **endodermis** separates the stele from the ground tissue in the outer portion of the root. The endodermis is exclusive to roots, and serves as a checkpoint for materials entering the root's vascular system. A waxy substance called suberin is present on the walls of the endodermal cells. This waxy region, known as the **Casparian strip**, forces water and solutes to cross the plasma membranes of endodermal cells instead of slipping between the cells. This ensures that only materials required by the root pass through the endodermis, while toxic substances and pathogens are generally excluded.

The outermost cell layer of the root's vascular tissue is the **pericycle**, an area that can give rise to lateral roots. In dicot roots, the xylem and phloem of the stele are arranged alternately in an X shape, whereas in monocot roots, the vascular tissue is arranged in a ring around the pith.



In (left) typical dicots, the vascular tissue forms an X shape in the center of the root. In (right) typical monocots, the phloem cells and the larger xylem cells form a characteristic ring around the central pith.

## Root Modifications

Root structures may be modified for specific purposes. For example, some roots are bulbous and store starch. Aerial roots and prop roots are two forms of aboveground roots that provide additional support to anchor the plant. Tap roots, such as carrots, turnips, and beets, are examples of roots that are modified for food storage ([link](#)).





Many vegetables are modified roots.

Epiphytic roots enable a plant to grow on another plant. For example, the epiphytic roots of orchids develop a spongy tissue to absorb moisture. The banyan tree (*Ficus* sp.) begins as an epiphyte, germinating in the branches of a host tree; aerial roots develop from the branches and eventually reach the ground, providing additional support ([link](#)). In screwpine (*Pandanus* sp.), a palm-like tree that grows in sandy tropical soils, aboveground prop roots develop from the nodes to provide additional support.



(a)



(b)

The (a) banyan tree, also known as the strangler fig, begins life as an epiphyte in a host tree. Aerial roots extend to the ground and support the growing plant,



which eventually strangles the host tree. The (b) screwpine develops aboveground roots that help support the plant in sandy soils. (credit a: modification of work by "psyberartist"/Flickr; credit b: modification of work by David Eikhoff)

## Section Summary

Roots help to anchor a plant, absorb water and minerals, and serve as storage sites for food. Taproots and fibrous roots are the two main types of root systems. In a taproot system, a main root grows vertically downward with a few lateral roots. Fibrous root systems arise at the base of the stem, where a cluster of roots forms a dense network that is shallower than a taproot. The growing root tip is protected by a root cap. The root tip has three main zones: a zone of cell division (cells are actively dividing), a zone of elongation (cells increase in length), and a zone of maturation (cells differentiate to form different kinds of cells). Root vascular tissue conducts water, minerals, and sugars. In some habitats, the roots of certain plants may be modified to form aerial roots or epiphytic roots.

## Review Questions

### Exercise:

#### Problem:

Roots that enable a plant to grow on another plant are called \_\_\_\_\_.

- a. epiphytic roots
- b. prop roots
- c. adventitious roots
- d. aerial roots

---

**Solution:**

A

**Exercise:**

**Problem:** The \_\_\_\_\_ forces selective uptake of minerals in the root.

- a. pericycle
- b. epidermis
- c. endodermis
- d. root cap

---

**Solution:**

C

**Exercise:**

**Problem:**

Newly-formed root cells begin to form different cell types in the \_\_\_\_\_.

- a. zone of elongation
- b. zone of maturation
- c. root meristem
- d. zone of cell division

---

**Solution:**

B

**Free Response**

**Exercise:**

**Problem:**

Compare a tap root system with a fibrous root system. For each type, name a plant that provides a food in the human diet. Which type of root system is found in monocots? Which type of root system is found in dicots?

---

**Solution:**

A tap root system has a single main root that grows down. A fibrous root system forms a dense network of roots that is closer to the soil surface. An example of a tap root system is a carrot. Grasses such as wheat, rice, and corn are examples of fibrous root systems. Fibrous root systems are found in monocots; tap root systems are found in dicots.

**Exercise:**

**Problem:** What might happen to a root if the pericycle disappeared?

---

**Solution:**

The root would not be able to produce lateral roots.

**Glossary**

adventitious root

aboveground root that arises from a plant part other than the radicle of the plant embryo

Casparian strip

waxy coating that forces water to cross endodermal plasma membranes before entering the vascular cylinder, instead of moving between endodermal cells

endodermis

layer of cells in the root that forms a selective barrier between the ground tissue and the vascular tissue, allowing water and minerals to enter the root while excluding toxins and pathogens

fibrous root system

type of root system in which the roots arise from the base of the stem in a cluster, forming a dense network of roots; found in monocots

pericycle

outer boundary of the stele from which lateral roots can arise

root cap

protective cells covering the tip of the growing root

root hair

hair-like structure that is an extension of epidermal cells; increases the root surface area and aids in absorption of water and minerals

stele

inner portion of the root containing the vascular tissue; surrounded by the endodermis

tap root system

type of root system with a main root that grows vertically with few lateral roots; found in dicots

## Leaves

By the end of this section, you will be able to:

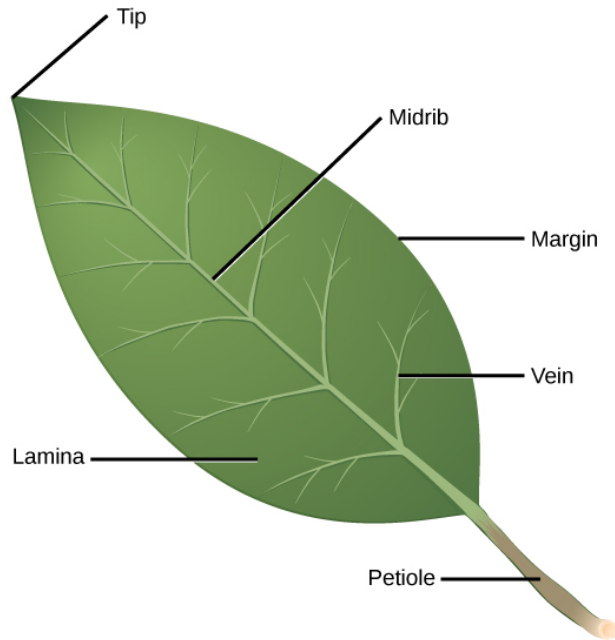
- Identify the parts of a typical leaf
- Describe the internal structure and function of a leaf
- Compare and contrast simple leaves and compound leaves
- List and describe examples of modified leaves

Leaves are the main sites for photosynthesis: the process by which plants synthesize food. Most leaves are usually green, due to the presence of chlorophyll in the leaf cells. However, some leaves may have different colors, caused by other plant pigments that mask the green chlorophyll.

The thickness, shape, and size of leaves are adapted to the environment. Each variation helps a plant species maximize its chances of survival in a particular habitat. Usually, the leaves of plants growing in tropical rainforests have larger surface areas than those of plants growing in deserts or very cold conditions, which are likely to have a smaller surface area to minimize water loss.

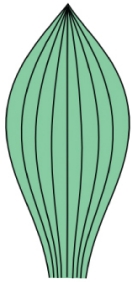
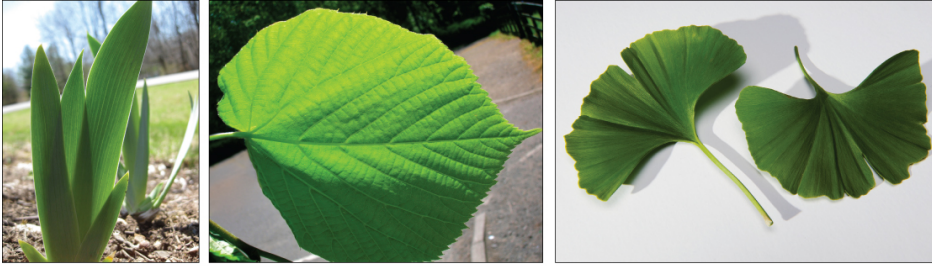
### Structure of a Typical Leaf

Each leaf typically has a leaf blade called the **lamina**, which is also the widest part of the leaf. Some leaves are attached to the plant stem by a **petiole**. Leaves that do not have a petiole and are directly attached to the plant stem are called **sessile** leaves. Small green appendages usually found at the base of the petiole are known as **stipules**. Most leaves have a midrib, which travels the length of the leaf and branches to each side to produce veins of vascular tissue. The edge of the leaf is called the margin. [\[link\]](#) shows the structure of a typical eudicot leaf.

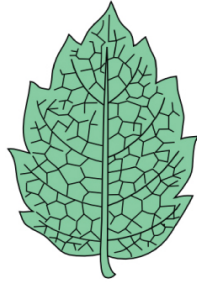


Deceptively simple in appearance,  
a leaf is a highly efficient  
structure.

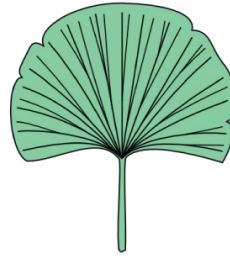
Within each leaf, the vascular tissue forms veins. The arrangement of veins in a leaf is called the **venation** pattern. Monocots and dicots differ in their patterns of venation ([\[link\]](#)). Monocots have parallel venation; the veins run in straight lines across the length of the leaf without converging at a point. In dicots, however, the veins of the leaf have a net-like appearance, forming a pattern known as reticulate venation. One extant plant, the *Ginkgo biloba*, has dichotomous venation where the veins fork.



(a)



(b)



(c)

(a) Tulip (*Tulipa*), a monocot, has leaves with parallel venation. The netlike venation in this (b) linden (*Tilia cordata*) leaf distinguishes it as a dicot.

The (c) *Ginkgo biloba* tree has dichotomous venation. (credit a photo: modification of work by “Drewboy64”/Wikimedia Commons; credit b photo: modification of work by Roger Griffith; credit c photo: modification of work by “geishaboy500”/Flickr; credit abc illustrations: modification of work by Agnieszka Kwiecień)

## Leaf Arrangement

The arrangement of leaves on a stem is known as **phyllotaxy**. The number and placement of a plant’s leaves will vary depending on the species, with each species exhibiting a characteristic leaf arrangement. Leaves are classified as either alternate, spiral, or opposite. Plants that have only one leaf per node have leaves that are said to be either alternate—meaning the leaves alternate on each side of the stem in a flat plane—or spiral, meaning the leaves are arrayed in a spiral along the stem. In an opposite leaf arrangement, two leaves arise at the same point, with the leaves connecting

opposite each other along the branch. If there are three or more leaves connected at a node, the leaf arrangement is classified as **whorled**.

## Leaf Form

Leaves may be simple or compound ([\[link\]](#)). In a **simple leaf**, the blade is either completely undivided—as in the banana leaf—or it has lobes, but the separation does not reach the midrib, as in the maple leaf. In a **compound leaf**, the leaf blade is completely divided, forming leaflets, as in the locust tree. Each leaflet may have its own stalk, but is attached to the rachis. A **palmately compound leaf** resembles the palm of a hand, with leaflets radiating outwards from one point. Examples include the leaves of poison ivy, the buckeye tree, or the familiar houseplant *Schefflera* sp. (common name “umbrella plant”). **Pinnately compound leaves** take their name from their feather-like appearance; the leaflets are arranged along the midrib, as in rose leaves (*Rosa* sp.), or the leaves of hickory, pecan, ash, or walnut trees.



(a) Simple



(b) Palmately compound



(c) Pinnately compound



(d) Doubly compound



Leaves may be simple or compound. In simple leaves, the lamina is continuous.

The (a) banana plant (*Musa* sp.) has simple leaves. In compound leaves, the lamina is separated into leaflets.

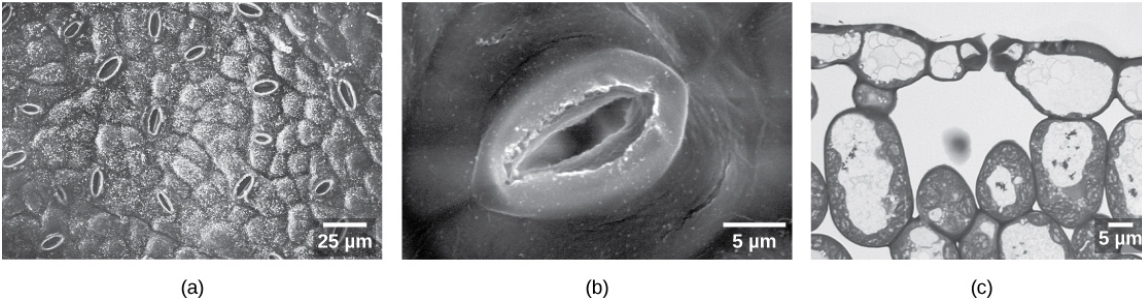
Compound leaves may be palmate or pinnate. In (b) palmately compound leaves, such as those of the horse chestnut (*Aesculus hippocastanum*), the leaflets branch from the petiole. In (c) pinnately compound leaves, the leaflets branch from the midrib, as on a scrub hickory (*Carya floridana*). The (d) honey locust has double compound leaves, in which leaflets branch from the veins.

(credit a: modification of work by "BazzaDaRambler"/Flickr; credit b: modification of work by Roberto Verzo; credit c: modification of work by Eric Dion; credit d: modification of work by Valerie Lykes)

## Leaf Structure and Function

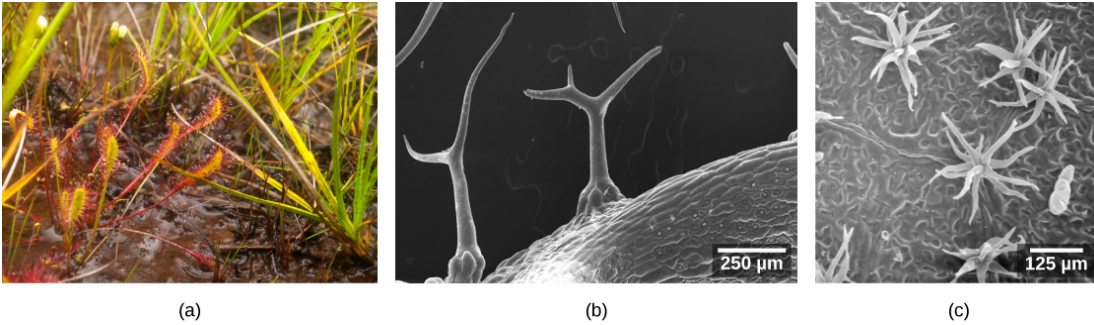
The outermost layer of the leaf is the epidermis; it is present on both sides of the leaf and is called the upper and lower epidermis, respectively.

Botanists call the upper side the adaxial surface (or adaxis) and the lower side the abaxial surface (or abaxis). The epidermis helps in the regulation of gas exchange. It contains stomata ([link](#)): openings through which the exchange of gases takes place. Two guard cells surround each stoma, regulating its opening and closing.



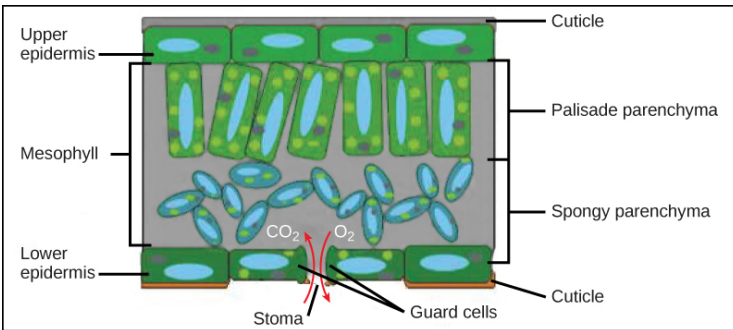
Visualized at 500x with a scanning electron microscope, several stomata are clearly visible on (a) the surface of this sumac (*Rhus glabra*) leaf. At 5,000x magnification, the guard cells of (b) a single stoma from lyre-leaved sand cress (*Arabidopsis lyrata*) have the appearance of lips that surround the opening. In this (c) light micrograph cross-section of an *A. lyrata* leaf, the guard cell pair is visible along with the large, sub-stomatal air space in the leaf. (credit: modification of work by Robert R. Wise; part c scale-bar data from Matt Russell)

The epidermis is usually one cell layer thick; however, in plants that grow in very hot or very cold conditions, the epidermis may be several layers thick to protect against excessive water loss from transpiration. A waxy layer known as the **cuticle** covers the leaves of all plant species. The cuticle reduces the rate of water loss from the leaf surface. Other leaves may have small hairs (trichomes) on the leaf surface. Trichomes help to deter herbivory by restricting insect movements, or by storing toxic or bad-tasting compounds; they can also reduce the rate of transpiration by blocking air flow across the leaf surface ([\[link\]](#)).

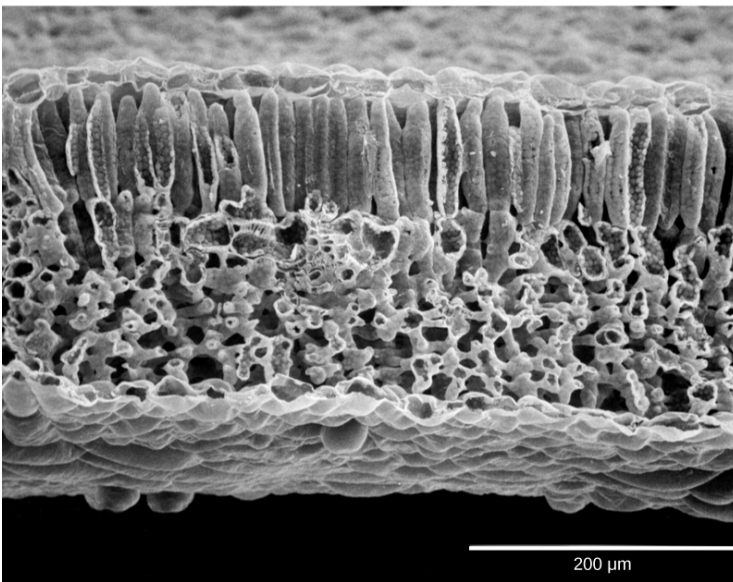


Trichomes give leaves a fuzzy appearance as in this (a) sundew (*Drosera* sp.). Leaf trichomes include (b) branched trichomes on the leaf of *Arabidopsis lyrata* and (c) multibranched trichomes on a mature *Quercus marilandica* leaf. (credit a: John Freeland; credit b, c: modification of work by Robert R. Wise; scale-bar data from Matt Russell)

Below the epidermis of dicot leaves are layers of cells known as the mesophyll, or “middle leaf.” The mesophyll of most leaves typically contains two arrangements of parenchyma cells: the palisade parenchyma and spongy parenchyma ([\[link\]](#)). The palisade parenchyma (also called the palisade mesophyll) has column-shaped, tightly packed cells, and may be present in one, two, or three layers. Below the palisade parenchyma are loosely arranged cells of an irregular shape. These are the cells of the spongy parenchyma (or spongy mesophyll). The air space found between the spongy parenchyma cells allows gaseous exchange between the leaf and the outside atmosphere through the stomata. In aquatic plants, the intercellular spaces in the spongy parenchyma help the leaf float. Both layers of the mesophyll contain many chloroplasts. Guard cells are the only epidermal cells to contain chloroplasts.



(a)

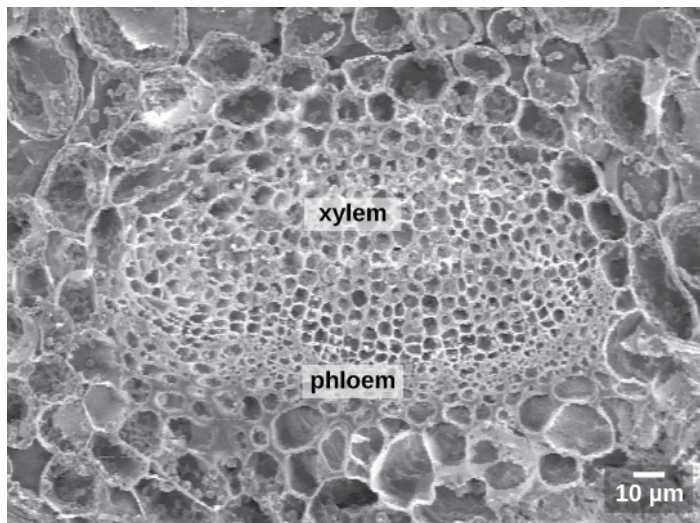


(b)

In the (a) leaf drawing, the central mesophyll is sandwiched between an upper and lower epidermis. The mesophyll has two layers: an upper palisade layer comprised of tightly packed, columnar cells, and a lower spongy layer, comprised of loosely packed, irregularly shaped cells. Stomata on the leaf underside allow gas exchange. A waxy cuticle covers all aerial surfaces of land plants to minimize water loss. These leaf layers are clearly visible in the (b) scanning electron micrograph. The numerous small bumps

in the palisade parenchyma cells are chloroplasts. Chloroplasts are also present in the spongy parenchyma, but are not as obvious. The bumps protruding from the lower surface of the leaf are glandular trichomes, which differ in structure from the stalked trichomes in [\[link\]](#). (credit b: modification of work by Robert R. Wise)

Like the stem, the leaf contains vascular bundles composed of xylem and phloem ([\[link\]](#)). The xylem consists of tracheids and vessels, which transport water and minerals to the leaves. The phloem transports the photosynthetic products from the leaf to the other parts of the plant. A single vascular bundle, no matter how large or small, always contains both xylem and phloem tissues.



This scanning electron micrograph shows xylem and phloem in the leaf vascular bundle from the lyre-leaved sand cress (*Arabidopsis lyrata*).

(credit: modification of work by  
Robert R. Wise; scale-bar data from  
Matt Russell)

## Leaf Adaptations

Coniferous plant species that thrive in cold environments, like spruce, fir, and pine, have leaves that are reduced in size and needle-like in appearance. These needle-like leaves have sunken stomata and a smaller surface area: two attributes that aid in reducing water loss. In hot climates, plants such as cacti have leaves that are reduced to spines, which in combination with their succulent stems, help to conserve water. Many aquatic plants have leaves with wide lamina that can float on the surface of the water, and a thick waxy cuticle on the leaf surface that repels water.

### Note:

Link to Learning



Watch “The Pale Pitcher Plant” episode of the [video](#) series *Plants Are Cool, Too*, a Botanical Society of America video about a carnivorous plant species found in Louisiana.

### Note:

Evolution Connection

**Plant Adaptations in Resource-Deficient Environments**



Roots, stems, and leaves are structured to ensure that a plant can obtain the required sunlight, water, soil nutrients, and oxygen resources. Some remarkable adaptations have evolved to enable plant species to thrive in less than ideal habitats, where one or more of these resources is in short supply.

In tropical rainforests, light is often scarce, since many trees and plants grow close together and block much of the sunlight from reaching the forest floor. Many tropical plant species have exceptionally broad leaves to maximize the capture of sunlight. Other species are epiphytes: plants that grow on other plants that serve as a physical support. Such plants are able to grow high up in the canopy atop the branches of other trees, where sunlight is more plentiful. Epiphytes live on rain and minerals collected in the branches and leaves of the supporting plant. Bromeliads (members of the pineapple family), ferns, and orchids are examples of tropical epiphytes ([link](#)). Many epiphytes have specialized tissues that enable them to efficiently capture and store water.



One of the most well known bromeliads is Spanish moss (*Tillandsia usneoides*), seen here in an oak tree. (credit: Kristine Paulus)

Some plants have special adaptations that help them to survive in nutrient-poor environments. Carnivorous plants, such as the Venus flytrap and the pitcher plant ([link](#)), grow in bogs where the soil is low in nitrogen. In these plants, leaves are modified to capture insects. The insect-capturing leaves may have evolved to provide these plants with a supplementary source of much-needed nitrogen.



(a)



(b)

The (a) Venus flytrap has modified leaves that can capture insects. When an unlucky insect touches the trigger hairs inside the leaf, the trap suddenly closes. The opening of the (b) pitcher plant is lined with a slippery wax. Insects crawling on the lip slip and fall into a pool of water in the bottom of the pitcher, where they are digested by bacteria.

The plant then absorbs the smaller molecules.

(credit a: modification of work by Peter Shanks;

credit b: modification of work by Tim Mansfield)

Many swamp plants have adaptations that enable them to thrive in wet areas, where their roots grow submerged underwater. In these aquatic areas, the soil is unstable and little oxygen is available to reach the roots. Trees such as mangroves (*Rhizophora* sp.) growing in coastal waters produce aboveground roots that help support the tree ([link](#)). Some species of mangroves, as well as cypress trees, have pneumatophores: upward-



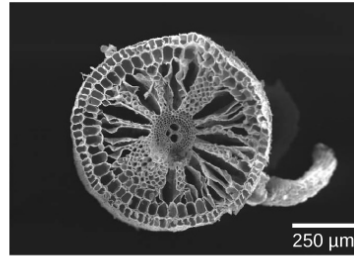
growing roots containing pores and pockets of tissue specialized for gas exchange. Wild rice is an aquatic plant with large air spaces in the root cortex. The air-filled tissue—called aerenchyma—provides a path for oxygen to diffuse down to the root tips, which are embedded in oxygen-poor bottom sediments.



(a)



(b)



(c)

The branches of (a) mangrove trees develop aerial roots, which descend to the ground and help to anchor the trees. (b) Cypress trees and some mangrove species have upward-growing roots called pneumatophores that are involved in gas exchange. Aquatic plants such as (c) wild rice have large spaces in the root cortex called aerenchyma, visualized here using scanning electron microscopy. (credit a: modification of work by Roberto Verzo; credit b: modification of work by Duane Burdick; credit c: modification of work by Robert R. Wise)

### Note:

#### Link to Learning



Watch *Venus Flytraps: Jaws of Death*, an extraordinary BBC close-up of the Venus flytrap in action.

## Section Summary

Leaves are the main site of photosynthesis. A typical leaf consists of a lamina (the broad part of the leaf, also called the blade) and a petiole (the stalk that attaches the leaf to a stem). The arrangement of leaves on a stem, known as phyllotaxy, enables maximum exposure to sunlight. Each plant species has a characteristic leaf arrangement and form. The pattern of leaf arrangement may be alternate, opposite, or spiral, while leaf form may be simple or compound. Leaf tissue consists of the epidermis, which forms the outermost cell layer, and mesophyll and vascular tissue, which make up the inner portion of the leaf. In some plant species, leaf form is modified to form structures such as tendrils, spines, bud scales, and needles.

## Review Questions

### Exercise:

**Problem:** The stalk of a leaf is known as the \_\_\_\_\_.

- a. petiole
- b. lamina
- c. stipule
- d. rachis

---

### Solution:

A

### Exercise:

**Problem:** Leaflets are a characteristic of \_\_\_\_\_ leaves.

- a. alternate
- b. whorled
- c. compound
- d. opposite

---

**Solution:**

C

**Exercise:**

**Problem:** Cells of the \_\_\_\_\_ contain chloroplasts.

- a. epidermis
- b. vascular tissue
- c. stomata
- d. mesophyll

---

**Solution:**

D

**Exercise:**

**Problem:**

Which of the following is most likely to be found in a desert environment?

- a. broad leaves to capture sunlight
- b. spines instead of leaves
- c. needle-like leaves
- d. wide, flat leaves that can float

---

**Solution:**

B

## Free Response

### Exercise:

#### Problem:

How do dicots differ from monocots in terms of leaf structure?

---

#### Solution:

Monocots have leaves with parallel venation, and dicots have leaves with reticulate, net-like venation.

### Exercise:

#### Problem:

Describe an example of a plant with leaves that are adapted to cold temperatures.

---

#### Solution:

Conifers such as spruce, fir, and pine have needle-shaped leaves with sunken stomata, helping to reduce water loss.

## Glossary

compound leaf

leaf in which the leaf blade is subdivided to form leaflets, all attached to the midrib

cuticle

waxy protective layer on the leaf surface

lamina

leaf blade

palmately compound leaf

leaf type with leaflets that emerge from a point, resembling the palm of a hand

petiole

stalk of the leaf

phyllotaxy

arrangement of leaves on a stem

pinnately compound leaf

leaf type with a divided leaf blade consisting of leaflets arranged on both sides of the midrib

sessile

leaf without a petiole that is attached directly to the plant stem

simple leaf

leaf type in which the lamina is completely undivided or merely lobed

stipule

small green structure found on either side of the leaf stalk or petiole

venation

pattern of veins in a leaf; may be parallel (as in monocots), reticulate (as in dicots), or dichotomous (as in *Ginkgo biloba*)

whorled

pattern of leaf arrangement in which three or more leaves are connected at a node

## Transport of Water and Solutes in Plants

By the end of this section, you will be able to:

- Define water potential and explain how it is influenced by solutes, pressure, gravity, and the matric potential
- Describe how water potential, evapotranspiration, and stomatal regulation influence how water is transported in plants
- Explain how photosynthates are transported in plants

The structure of plant roots, stems, and leaves facilitates the transport of water, nutrients, and photosynthates throughout the plant. The phloem and xylem are the main tissues responsible for this movement. Water potential, evapotranspiration, and stomatal regulation influence how water and nutrients are transported in plants. To understand how these processes work, we must first understand the energetics of water potential.

### Water Potential

Plants are phenomenal hydraulic engineers. Using only the basic laws of physics and the simple manipulation of potential energy, plants can move water to the top of a 116-meter-tall tree ([link](#)a). Plants can also use hydraulics to generate enough force to split rocks and buckle sidewalks ([link](#)b). Plants achieve this because of water potential.



(a)



(b)

With heights nearing 116 meters, (a) coastal redwoods (*Sequoia sempervirens*) are the tallest trees in the world. Plant roots can easily generate enough force to (b) buckle and break concrete sidewalks, much to the dismay of homeowners and city maintenance departments. (credit a: modification of work by Bernt Rostad; credit b: modification of work by Pedestrians Educating Drivers on Safety, Inc.)

**Water potential** is a measure of the potential energy in water. Plant physiologists are not interested in the energy in any one particular aqueous system, but are very interested in water movement between two systems. In practical terms, therefore, water potential is the difference in potential energy between a given water sample and pure water (at atmospheric pressure and ambient temperature). Water potential is denoted by the Greek letter  $\psi$  (*psi*) and is expressed in units of pressure (pressure is a form of energy) called **megapascals** (MPa). The potential of pure water ( $\Psi_w^{\text{pure H}_2\text{O}}$ ) is, by convenience of definition, designated a value of zero (even though pure water contains plenty of potential energy, that energy is ignored). Water potential values for the water in a plant root, stem, or leaf are therefore expressed relative to  $\Psi_w^{\text{pure H}_2\text{O}}$ .

The water potential in plant solutions is influenced by solute concentration, pressure, gravity, and factors called matrix effects. Water potential can be broken down into its individual components using the following equation:

**Equation:**

$$\Psi_{\text{system}} = \Psi_{\text{total}} = \Psi_s + \Psi_p + \Psi_g + \Psi_m$$

where  $\Psi_s$ ,  $\Psi_p$ ,  $\Psi_g$ , and  $\Psi_m$  refer to the solute, pressure, gravity, and matric potentials, respectively. “System” can refer to the water potential of the soil water ( $\Psi^{\text{soil}}$ ), root water ( $\Psi^{\text{root}}$ ), stem water ( $\Psi^{\text{stem}}$ ), leaf water ( $\Psi^{\text{leaf}}$ ) or the water in the atmosphere ( $\Psi^{\text{atmosphere}}$ ): whichever aqueous system is under consideration. As the individual components change, they raise or lower the

total water potential of a system. When this happens, water moves to equilibrate, moving from the system or compartment with a higher water potential to the system or compartment with a lower water potential. This brings the difference in water potential between the two systems ( $\Delta\Psi$ ) back to zero ( $\Delta\Psi = 0$ ). Therefore, for water to move through the plant from the soil to the air (a process called transpiration),  $\Psi^{\text{soil}}$  must be  $> \Psi^{\text{root}} > \Psi^{\text{stem}} > \Psi^{\text{leaf}} > \Psi^{\text{atmosphere}}$ .

Water only moves in response to  $\Delta\Psi$ , not in response to the individual components. However, because the individual components influence the total  $\Psi_{\text{system}}$ , by manipulating the individual components (especially  $\Psi_s$ ), a plant can control water movement.

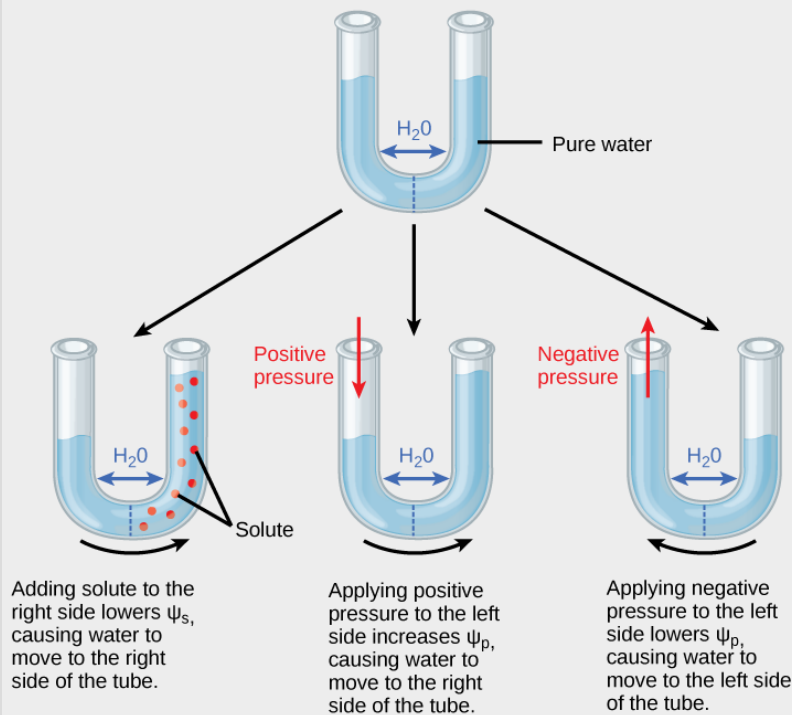
## Solute Potential

Solute potential ( $\Psi_s$ ), also called osmotic potential, is negative in a plant cell and zero in distilled water. Typical values for cell cytoplasm are  $-0.5$  to  $-1.0$  MPa. Solutes reduce water potential (resulting in a negative  $\Psi_w$ ) by consuming some of the potential energy available in the water. Solute molecules can dissolve in water because water molecules can bind to them via hydrogen bonds; a hydrophobic molecule like oil, which cannot bind to water, cannot go into solution. The energy in the hydrogen bonds between solute molecules and water is no longer available to do work in the system because it is tied up in the bond. In other words, the amount of available potential energy is reduced when solutes are added to an aqueous system. Thus,  $\Psi_s$  decreases with increasing solute concentration. Because  $\Psi_s$  is one of the four components of  $\Psi_{\text{system}}$  or  $\Psi_{\text{total}}$ , a decrease in  $\Psi_s$  will cause a decrease in  $\Psi_{\text{total}}$ . The internal water potential of a plant cell is more negative than pure water because of the cytoplasm's high solute content ([\[link\]](#)). Because of this difference in water potential water will move from the soil into a plant's root cells via the process of osmosis. This is why solute potential is sometimes called osmotic potential.

Plant cells can metabolically manipulate  $\Psi_s$  (and by extension,  $\Psi_{\text{total}}$ ) by adding or removing solute molecules. Therefore, plants have control over  $\Psi_{\text{total}}$  via their ability to exert metabolic control over  $\Psi_s$ .



**Note:**



In this example with a semipermeable membrane between two aqueous systems, water will move from a region of higher to lower water potential until equilibrium is reached. Solutes ( $\Psi_s$ ), pressure ( $\Psi_p$ ), and gravity ( $\Psi_g$ ) influence total water potential for each side of the tube ( $\Psi_{\text{total}}^{\text{right or left}}$ ), and therefore, the difference between  $\Psi_{\text{total}}$  on each side ( $\Delta\Psi$ ). ( $\Psi_m$ , the potential due to interaction of water with solid substrates, is ignored in this example because glass is not especially hydrophilic). Water moves in response to the difference in water potential between two systems (the left and right sides of the tube).

Positive water potential is placed on the left side of the tube by increasing  $\Psi_p$  such that the water level rises on the right side. Could you equalize the

water level on each side of the tube by adding solute, and if so, how?

## Pressure Potential

Pressure potential ( $\Psi_p$ ), also called turgor potential, may be positive or negative ([\[link\]](#)). Because pressure is an expression of energy, the higher the pressure, the more potential energy in a system, and vice versa.

Therefore, a positive  $\Psi_p$  (compression) increases  $\Psi_{\text{total}}$ , and a negative  $\Psi_p$  (tension) decreases  $\Psi_{\text{total}}$ . Positive pressure inside cells is contained by the cell wall, producing turgor pressure. Pressure potentials are typically around 0.6–0.8 MPa, but can reach as high as 1.5 MPa in a well-watered plant. A  $\Psi_p$  of 1.5 MPa equates to 210 pounds per square inch ( $1.5 \text{ MPa} \times 140 \text{ lb in}^{-2} \text{ MPa}^{-1} = 210 \text{ lb/in}^{-2}$ ). As a comparison, most automobile tires are kept at a pressure of 30–34 psi. An example of the effect of turgor pressure is the wilting of leaves and their restoration after the plant has been watered ([\[link\]](#)). Water is lost from the leaves via transpiration (approaching  $\Psi_p = 0$  MPa at the wilting point) and restored by uptake via the roots.

A plant can manipulate  $\Psi_p$  via its ability to manipulate  $\Psi_s$  and by the process of osmosis. If a plant cell increases the cytoplasmic solute concentration,  $\Psi_s$  will decline,  $\Psi_{\text{total}}$  will decline, the  $\Delta\Psi$  between the cell and the surrounding tissue will decline, water will move into the cell by osmosis, and  $\Psi_p$  will increase.  $\Psi_p$  is also under indirect plant control via the opening and closing of stomata. Stomatal openings allow water to evaporate from the leaf, reducing  $\Psi_p$  and  $\Psi_{\text{total}}$  of the leaf and increasing it between the water in the leaf and the petiole, thereby allowing water to flow from the petiole into the leaf.



When (a) total water potential ( $\Psi_{\text{total}}$ ) is lower outside the cells than inside, water moves out of the cells and the plant wilts. When (b) the total water potential is higher outside the plant cells than inside, water moves into the cells, resulting in turgor pressure ( $\Psi_p$ ) and keeping the plant erect. (credit: modification of work by Victor M. Vicente Selvas)

## Gravity Potential

Gravity potential ( $\Psi_g$ ) is always negative to zero in a plant with no height. It always removes or consumes potential energy from the system. The force of gravity pulls water downwards to the soil, reducing the total amount of potential energy in the water in the plant ( $\Psi_{\text{total}}$ ). The taller the plant, the taller the water column, and the more influential  $\Psi_g$  becomes. On a cellular scale and in short plants, this effect is negligible and easily ignored. However, over the height of a tall tree like a giant coastal redwood, the gravitational pull of  $-0.1 \text{ MPa m}^{-1}$  is equivalent to an extra 1 MPa of resistance that must be overcome for water to reach the leaves of the tallest trees. Plants are unable to manipulate  $\Psi_g$ .

## Matric Potential

Matric potential ( $\Psi_m$ ) is always negative to zero. In a dry system, it can be as low as  $-2$  MPa in a dry seed, and it is zero in a water-saturated system. The binding of water to a matrix always removes or consumes potential energy from the system.  $\Psi_m$  is similar to solute potential because it involves tying up the energy in an aqueous system by forming hydrogen bonds between the water and some other component. However, in solute potential, the other components are soluble, hydrophilic solute molecules, whereas in  $\Psi_m$ , the other components are insoluble, hydrophilic molecules of the plant cell wall. Every plant cell has a cellulosic cell wall and the cellulose in the cell walls is hydrophilic, producing a matrix for adhesion of water: hence the name matric potential.  $\Psi_m$  is very large (negative) in dry tissues such as seeds or drought-affected soils. However, it quickly goes to zero as the seed takes up water or the soil hydrates.  $\Psi_m$  cannot be manipulated by the plant and is typically ignored in well-watered roots, stems, and leaves.

## Movement of Water and Minerals in the Xylem

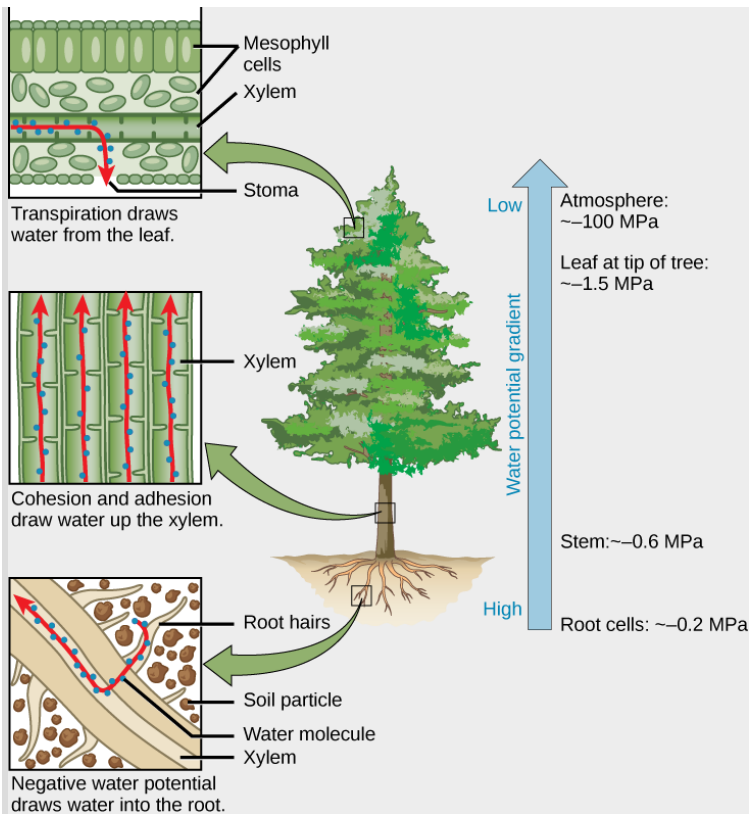
Solutes, pressure, gravity, and matric potential are all important for the transport of water in plants. Water moves from an area of higher total water potential (higher Gibbs free energy) to an area of lower total water potential. Gibbs free energy is the energy associated with a chemical reaction that can be used to do work. This is expressed as  $\Delta\Psi$ .

**Transpiration** is the loss of water from the plant through evaporation at the leaf surface. It is the main driver of water movement in the xylem. Transpiration is caused by the evaporation of water at the leaf–atmosphere interface; it creates negative pressure (tension) equivalent to  $-2$  MPa at the leaf surface. This value varies greatly depending on the vapor pressure deficit, which can be negligible at high relative humidity (RH) and substantial at low RH. Water from the roots is pulled up by this tension. At night, when stomata shut and transpiration stops, the water is held in the stem and leaf by the adhesion of water to the cell walls of the xylem vessels and tracheids, and the cohesion of water molecules to each other. This is called the cohesion–tension theory of sap ascent.

Inside the leaf at the cellular level, water on the surface of mesophyll cells saturates the cellulose microfibrils of the primary cell wall. The leaf contains many large intercellular air spaces for the exchange of oxygen for carbon dioxide, which is required for photosynthesis. The wet cell wall is exposed to this leaf internal air space, and the water on the surface of the cells evaporates into the air spaces, decreasing the thin film on the surface of the mesophyll cells. This decrease creates a greater tension on the water in the mesophyll cells ([\[link\]](#)), thereby increasing the pull on the water in the xylem vessels. The xylem vessels and tracheids are structurally adapted to cope with large changes in pressure. Rings in the vessels maintain their tubular shape, much like the rings on a vacuum cleaner hose keep the hose open while it is under pressure. Small perforations between vessel elements reduce the number and size of gas bubbles that can form via a process called cavitation. The formation of gas bubbles in xylem interrupts the continuous stream of water from the base to the top of the plant, causing a break termed an embolism in the flow of xylem sap. The taller the tree, the greater the tension forces needed to pull water, and the more cavitation events. In larger trees, the resulting embolisms can plug xylem vessels, making them non-functional.

**Note:**

Art Connection



The cohesion–tension theory of sap ascent is shown. Evaporation from the mesophyll cells produces a negative water potential gradient that causes water to move upwards from the roots through the xylem.

Which of the following statements is false?

- Negative water potential draws water into the root hairs. Cohesion and adhesion draw water up the xylem. Transpiration draws water from the leaf.
- Negative water potential draws water into the root hairs. Cohesion and adhesion draw water up the phloem. Transpiration draws water from the leaf.
- Water potential decreases from the roots to the top of the plant.
- Water enters the plants through root hairs and exits through stoma.

---

**Transpiration**—the loss of water vapor to the atmosphere through stomata—is a passive process, meaning that metabolic energy in the form of ATP is not required for water movement. The energy driving transpiration is the difference in energy between the water in the soil and the water in the atmosphere. However, transpiration is tightly controlled.

## Control of Transpiration

The atmosphere to which the leaf is exposed drives transpiration, but also causes massive water loss from the plant. Up to 90 percent of the water taken up by roots may be lost through transpiration.

Leaves are covered by a waxy **cuticle** on the outer surface that prevents the loss of water. Regulation of transpiration, therefore, is achieved primarily through the opening and closing of stomata on the leaf surface. Stomata are surrounded by two specialized cells called guard cells, which open and close in response to environmental cues such as light intensity and quality, leaf water status, and carbon dioxide concentrations. Stomata must open to allow air containing carbon dioxide and oxygen to diffuse into the leaf for photosynthesis and respiration. When stomata are open, however, water vapor is lost to the external environment, increasing the rate of transpiration. Therefore, plants must maintain a balance between efficient photosynthesis and water loss.

Plants have evolved over time to adapt to their local environment and reduce transpiration([link](#)). Desert plant (xerophytes) and plants that grow on other plants (epiphytes) have limited access to water. Such plants usually have a much thicker waxy cuticle than those growing in more moderate, well-watered environments (mesophytes). Aquatic plants (hydrophytes) also have their own set of anatomical and morphological leaf adaptations.





(a)



(b)



(c)



(d)

Plants are suited to their local environment. (a) Xerophytes, like this prickly pear cactus (*Opuntia* sp.) and (b) epiphytes such as this tropical *Aeschynanthus perrottetii* have adapted to very limited water resources. The leaves of a prickly pear are modified into spines, which lowers the surface-to-volume ratio and reduces water loss. Photosynthesis takes place in the stem, which also stores water. (b) *A. perrottetii* leaves have a waxy cuticle that prevents water loss. (c) Goldenrod (*Solidago* sp.) is a mesophyte, well suited for moderate environments. (d)



Hydrophytes, like this fragrant water lily (*Nymphaea odorata*), are adapted to thrive in aquatic environments. (credit a: modification of work by Jon Sullivan; credit b: modification of work by L. Shyamal/Wikimedia Commons; credit c: modification of work by Huw Williams; credit d: modification of work by Jason Hollinger)

Xerophytes and epiphytes often have a thick covering of trichomes or of stomata that are sunken below the leaf's surface. Trichomes are specialized hair-like epidermal cells that secrete oils and substances. These adaptations impede air flow across the stomatal pore and reduce transpiration. Multiple epidermal layers are also commonly found in these types of plants.

## Transportation of Photosynthates in the Phloem

Plants need an energy source to grow. In seeds and bulbs, food is stored in polymers (such as starch) that are converted by metabolic processes into sucrose for newly developing plants. Once green shoots and leaves are growing, plants are able to produce their own food by photosynthesizing. The products of photosynthesis are called photosynthates, which are usually in the form of simple sugars such as sucrose.

Structures that produce photosynthates for the growing plant are referred to as **sources**. Sugars produced in sources, such as leaves, need to be delivered to growing parts of the plant via the phloem in a process called **translocation**. The points of sugar delivery, such as roots, young shoots, and developing seeds, are called **sinks**. Seeds, tubers, and bulbs can be either a source or a sink, depending on the plant's stage of development and the season.

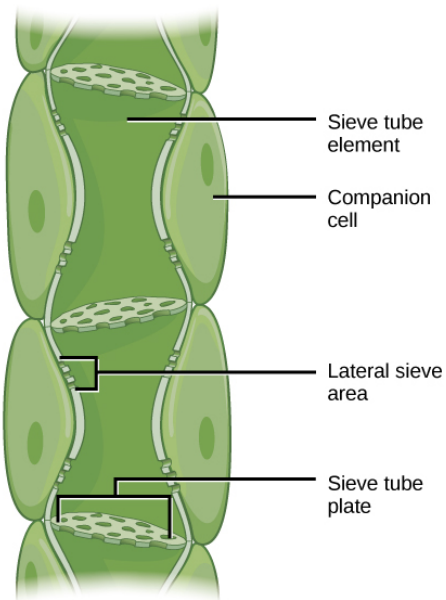
The products from the source are usually translocated to the nearest sink through the phloem. For example, the highest leaves will send

photosynthates upward to the growing shoot tip, whereas lower leaves will direct photosynthates downward to the roots. Intermediate leaves will send products in both directions, unlike the flow in the xylem, which is always unidirectional (soil to leaf to atmosphere). The pattern of photosynthate flow changes as the plant grows and develops. Photosynthates are directed primarily to the roots early on, to shoots and leaves during vegetative growth, and to seeds and fruits during reproductive development. They are also directed to tubers for storage.

### **Translocation: Transport from Source to Sink**

Photosynthates, such as sucrose, are produced in the mesophyll cells of photosynthesizing leaves. From there they are translocated through the phloem to where they are used or stored. Mesophyll cells are connected by cytoplasmic channels called plasmodesmata. Photosynthates move through these channels to reach phloem sieve-tube elements (STEs) in the vascular bundles. From the mesophyll cells, the photosynthates are loaded into the phloem STEs. The sucrose is actively transported against its concentration gradient (a process requiring ATP) into the phloem cells using the electrochemical potential of the proton gradient. This is coupled to the uptake of sucrose with a carrier protein called the sucrose- $H^+$  symporter.

Phloem STEs have reduced cytoplasmic contents, and are connected by a sieve plate with pores that allow for pressure-driven bulk flow, or translocation, of phloem sap. Companion cells are associated with STEs. They assist with metabolic activities and produce energy for the STEs ([link](#)).

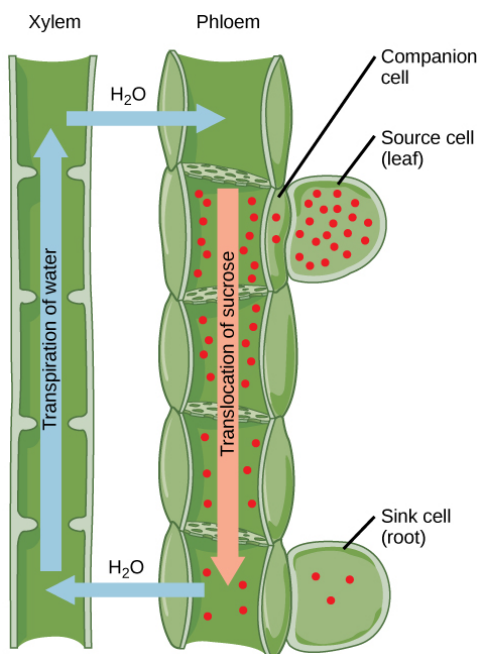


Phloem is comprised of cells called sieve-tube elements. Phloem sap travels through perforations called sieve tube plates.

Neighboring companion cells carry out metabolic functions for the sieve-tube elements and provide them with energy. Lateral sieve areas connect the sieve-tube elements to the companion cells.

Once in the phloem, the photosynthates are translocated to the closest sink. Phloem sap is an aqueous solution that contains up to 30 percent sugar, minerals, amino acids, and plant growth regulators. The high percentage of sugar decreases  $\Psi_s$ , which decreases the total water potential and causes

water to move by osmosis from the adjacent xylem into the phloem tubes, thereby increasing pressure. This increase in total water potential causes the bulk flow of phloem from source to sink ([link](#)). Sucrose concentration in the sink cells is lower than in the phloem STEs because the sink sucrose has been metabolized for growth, or converted to starch for storage or other polymers, such as cellulose, for structural integrity. Unloading at the sink end of the phloem tube occurs by either diffusion or active transport of sucrose molecules from an area of high concentration to one of low concentration. Water diffuses from the phloem by osmosis and is then transpired or recycled via the xylem back into the phloem sap.



Sucrose is actively transported from source cells into companion cells and then into the sieve-tube elements. This reduces the water potential, which causes water to enter the phloem from the xylem.

The resulting positive pressure forces the sucrose-water mixture down toward the roots, where sucrose is unloaded. Transpiration causes water to return to the leaves through the xylem vessels.

## Section Summary

Water potential ( $\Psi$ ) is a measure of the difference in potential energy between a water sample and pure water. The water potential in plant solutions is influenced by solute concentration, pressure, gravity, and matric potential. Water potential and transpiration influence how water is transported through the xylem in plants. These processes are regulated by stomatal opening and closing. Photosynthates (mainly sucrose) move from sources to sinks through the plant's phloem. Sucrose is actively loaded into the sieve-tube elements of the phloem. The increased solute concentration causes water to move by osmosis from the xylem into the phloem. The positive pressure that is produced pushes water and solutes down the pressure gradient. The sucrose is unloaded into the sink, and the water returns to the xylem vessels.

## Art Connections

### Exercise:

#### Problem:

[\[link\]](#) Positive water potential is placed on the left side of the tube by increasing  $\Psi_p$  such that the water level rises on the right side. Could you equalize the water level on each side of the tube by adding solute, and if so, how?

---

**Solution:**

[\[link\]](#) Yes, you can equalize the water level by adding the solute to the left side of the tube such that water moves toward the left until the water levels are equal.

**Exercise:**

**Problem:** [\[link\]](#) Which of the following statements is false?

- a. Negative water potential draws water into the root hairs. Cohesion and adhesion draw water up the xylem. Transpiration draws water from the leaf.
- b. Negative water potential draws water into the root hairs. Cohesion and adhesion draw water up the phloem. Transpiration draws water from the leaf.
- c. Water potential decreases from the roots to the top of the plant.
- d. Water enters the plants through root hairs and exits through stoma.

---

**Solution:**

[\[link\]](#) B.

**Review Questions****Exercise:**

**Problem:** When stomata open, what occurs?

- a. Water vapor is lost to the external environment, increasing the rate of transpiration.
- b. Water vapor is lost to the external environment, decreasing the rate of transpiration.

- c. Water vapor enters the spaces in the mesophyll, increasing the rate of transpiration.
  - d. Water vapor enters the spaces in the mesophyll, increasing the rate of transpiration.
- 

**Solution:**

A

**Exercise:**

**Problem:**

Which cells are responsible for the movement of photosynthates through a plant?

- a. tracheids, vessel elements
  - b. tracheids, companion cells
  - c. vessel elements, companion cells
  - d. sieve-tube elements, companion cells
- 

**Solution:**

D

**Free Response**

**Exercise:**

**Problem:**

The process of bulk flow transports fluids in a plant. Describe the two main bulk flow processes.

---

**Solution:**

The process of bulk flow moves water up the xylem and moves photosynthates (solutes) up and down the phloem.

## **Glossary**

cuticle

waxy covering on the outside of the leaf and stem that prevents the loss of water

megapascal (MPa)

pressure units that measure water potential

sink

growing parts of a plant, such as roots and young leaves, which require photosynthate

source

organ that produces photosynthate for a plant

translocation

mass transport of photosynthates from source to sink in vascular plants

transpiration

loss of water vapor to the atmosphere through stomata

water potential ( $\Psi_w$ )

the potential energy of a water solution per unit volume in relation to pure water at atmospheric pressure and ambient temperature



## Plant Sensory Systems and Responses

By the end of this section, you will be able to:

- Describe how red and blue light affect plant growth and metabolic activities
- Discuss gravitropism
- Understand how hormones affect plant growth and development
- Describe thigmotropism, thigmonastism, and thigmogenesis
- Explain how plants defend themselves from predators and respond to wounds

Animals can respond to environmental factors by moving to a new location. Plants, however, are rooted in place and must respond to the surrounding environmental factors. Plants have sophisticated systems to detect and respond to light, gravity, temperature, and physical touch. Receptors sense environmental factors and relay the information to effector systems—often through intermediate chemical messengers—to bring about plant responses.

## Plant Responses to Light

Plants have a number of sophisticated uses for light that go far beyond their ability to photosynthesize low-molecular-weight sugars using only carbon dioxide, light, and water. **Photomorphogenesis** is the growth and development of plants in response to light. It allows plants to optimize their use of light and space. **Photoperiodism** is the ability to use light to track time. Plants can tell the time of day and time of year by sensing and using various wavelengths of sunlight. **Phototropism** is a directional response that allows plants to grow towards, or even away from, light.

The sensing of light in the environment is important to plants; it can be crucial for competition and survival. The response of plants to light is mediated by different photoreceptors, which are comprised of a protein covalently bonded to a light-absorbing pigment called a **chromophore**. Together, the two are called a chromoprotein.

The red/far-red and violet-blue regions of the visible light spectrum trigger structural development in plants. Sensory photoreceptors absorb light in

these particular regions of the visible light spectrum because of the quality of light available in the daylight spectrum. In terrestrial habitats, light absorption by chlorophylls peaks in the blue and red regions of the spectrum. As light filters through the canopy and the blue and red wavelengths are absorbed, the spectrum shifts to the far-red end, shifting the plant community to those plants better adapted to respond to far-red light. Blue-light receptors allow plants to gauge the direction and abundance of sunlight, which is rich in blue-green emissions. Water absorbs red light, which makes the detection of blue light essential for algae and aquatic plants.

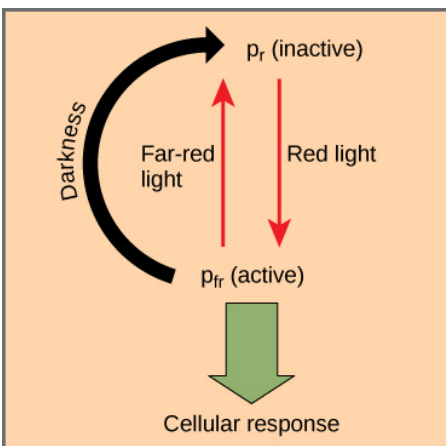
## **The Phytochrome System and the Red/Far-Red Response**

The **phytochromes** are a family of chromoproteins with a linear tetrapyrrole chromophore, similar to the ringed tetrapyrrole light-absorbing head group of chlorophyll. Phytochromes have two photo-interconvertible forms: Pr and Pfr. Pr absorbs red light (~667 nm) and is immediately converted to Pfr. Pfr absorbs far-red light (~730 nm) and is quickly converted back to Pr. Absorption of red or far-red light causes a massive change to the shape of the chromophore, altering the conformation and activity of the phytochrome protein to which it is bound. Pfr is the physiologically active form of the protein; therefore, exposure to red light yields physiological activity. Exposure to far-red light inhibits phytochrome activity. Together, the two forms represent the phytochrome system ([\[link\]](#)).

The phytochrome system acts as a biological light switch. It monitors the level, intensity, duration, and color of environmental light. The effect of red light is reversible by immediately shining far-red light on the sample, which converts the chromoprotein to the inactive Pr form. Additionally, Pfr can slowly revert to Pr in the dark, or break down over time. In all instances, the physiological response induced by red light is reversed. The active form of phytochrome (Pfr) can directly activate other molecules in the cytoplasm, or it can be trafficked to the nucleus, where it directly activates or represses specific gene expression.

Once the phytochrome system evolved, plants adapted it to serve a variety of needs. Unfiltered, full sunlight contains much more red light than far-red light. Because chlorophyll absorbs strongly in the red region of the visible spectrum, but not in the far-red region, any plant in the shade of another plant on the forest floor will be exposed to red-depleted, far-red-enriched light. The preponderance of far-red light converts phytochrome in the shaded leaves to the Pr (inactive) form, slowing growth. The nearest non-shaded (or even less-shaded) areas on the forest floor have more red light; leaves exposed to these areas sense the red light, which activates the Pfr form and induces growth. In short, plant shoots use the phytochrome system to grow away from shade and towards light. Because competition for light is so fierce in a dense plant community, the evolutionary advantages of the phytochrome system are obvious.

In seeds, the phytochrome system is not used to determine direction and quality of light (shaded versus unshaded). Instead, is it used merely to determine if there is any light at all. This is especially important in species with very small seeds, such as lettuce. Because of their size, lettuce seeds have few food reserves. Their seedlings cannot grow for long before they run out of fuel. If they germinated even a centimeter under the soil surface, the seedling would never make it into the sunlight and would die. In the dark, phytochrome is in the Pr (inactive form) and the seed will not germinate; it will only germinate if exposed to light at the surface of the soil. Upon exposure to light, Pr is converted to Pfr and germination proceeds.



The biologically inactive form of phytochrome (Pr) is converted to the biologically active form Pfr under illumination with red light. Far-red light and darkness convert the molecule back to the inactive form.

Plants also use the phytochrome system to sense the change of season. Photoperiodism is a biological response to the timing and duration of day and night. It controls flowering, setting of winter buds, and vegetative growth. Detection of seasonal changes is crucial to plant survival. Although temperature and light intensity influence plant growth, they are not reliable indicators of season because they may vary from one year to the next. Day length is a better indicator of the time of year.

As stated above, unfiltered sunlight is rich in red light but deficient in far-red light. Therefore, at dawn, all the phytochrome molecules in a leaf quickly convert to the active Pfr form, and remain in that form until sunset. In the dark, the Pfr form takes hours to slowly revert back to the Pr form. If the night is long (as in winter), all of the Pfr form reverts. If the night is short (as in summer), a considerable amount of Pfr may remain at sunrise. By sensing the Pr/Pfr ratio at dawn, a plant can determine the length of the day/night cycle. In addition, leaves retain that information for several days, allowing a comparison between the length of the previous night and the preceding several nights. Shorter nights indicate springtime to the plant; when the nights become longer, autumn is approaching. This information, along with sensing temperature and water availability, allows plants to determine the time of the year and adjust their physiology accordingly. Short-day (long-night) plants use this information to flower in the late

summer and early fall, when nights exceed a critical length (often eight or fewer hours). Long-day (short-night) plants flower during the spring, when darkness is less than a critical length (often eight to 15 hours). Not all plants use the phytochrome system in this way. Flowering in day-neutral plants is not regulated by daylength.

**Note:**

**Career Connection**

**Horticulturalist**

The word “horticulturist” comes from the Latin words for garden (*hortus*) and culture (*cultura*). This career has been revolutionized by progress made in the understanding of plant responses to environmental stimuli.

Growers of crops, fruit, vegetables, and flowers were previously constrained by having to time their sowing and harvesting according to the season. Now, horticulturists can manipulate plants to increase leaf, flower, or fruit production by understanding how environmental factors affect plant growth and development.

Greenhouse management is an essential component of a horticulturist’s education. To lengthen the night, plants are covered with a blackout shade cloth. Long-day plants are irradiated with red light in winter to promote early flowering. For example, fluorescent (cool white) light high in blue wavelengths encourages leafy growth and is excellent for starting seedlings. Incandescent lamps (standard light bulbs) are rich in red light, and promote flowering in some plants. The timing of fruit ripening can be increased or delayed by applying plant hormones. Recently, considerable progress has been made in the development of plant breeds that are suited to different climates and resistant to pests and transportation damage. Both crop yield and quality have increased as a result of practical applications of the knowledge of plant responses to external stimuli and hormones.

Horticulturists find employment in private and governmental laboratories, greenhouses, botanical gardens, and in the production or research fields. They improve crops by applying their knowledge of genetics and plant physiology. To prepare for a horticulture career, students take classes in botany, plant physiology, plant pathology, landscape design, and plant

breeding. To complement these traditional courses, horticulture majors add studies in economics, business, computer science, and communications.

## The Blue Light Responses

Phototropism—the directional bending of a plant toward or away from a light source—is a response to blue wavelengths of light. Positive phototropism is growth towards a light source ([link](#)), while negative phototropism (also called skototropism) is growth away from light.

The aptly-named **phototropins** are protein-based receptors responsible for mediating the phototropic response. Like all plant photoreceptors, phototropins consist of a protein portion and a light-absorbing portion, called the chromophore. In phototropins, the chromophore is a covalently-bound molecule of flavin; hence, phototropins belong to a class of proteins called flavoproteins.

Other responses under the control of phototropins are leaf opening and closing, chloroplast movement, and the opening of stomata. However, of all responses controlled by phototropins, phototropism has been studied the longest and is the best understood.

In their 1880 treatise *The Power of Movements in Plants*, Charles Darwin and his son Francis first described phototropism as the bending of seedlings toward light. Darwin observed that light was perceived by the tip of the plant (the apical meristem), but that the response (bending) took place in a different part of the plant. They concluded that the signal had to travel from the apical meristem to the base of the plant.



Azure bluets (*Houstonia caerulea*) display a phototropic response by bending toward the light. (credit: Cory Zanker)

In 1913, Peter Boysen-Jensen demonstrated that a chemical signal produced in the plant tip was responsible for the bending at the base. He cut off the tip of a seedling, covered the cut section with a layer of gelatin, and then replaced the tip. The seedling bent toward the light when illuminated. However, when impermeable mica flakes were inserted between the tip and the cut base, the seedling did not bend. A refinement of the experiment showed that the signal traveled on the shaded side of the seedling. When the mica plate was inserted on the illuminated side, the plant did bend towards the light. Therefore, the chemical signal was a growth stimulant because the phototropic response involved faster cell elongation on the shaded side than on the illuminated side. We now know that as light passes through a plant stem, it is diffracted and generates phototropin activation across the stem. Most activation occurs on the lit side, causing the plant hormone indole acetic acid (IAA) to accumulate on the shaded side. Stem cells elongate under influence of IAA.

**Cryptochromes** are another class of blue-light absorbing photoreceptors that also contain a flavin-based chromophore. Cryptochromes set the plants

24-hour activity cycle, also known as its circadian rhythm, using blue light cues. There is some evidence that cryptochromes work together with phototropins to mediate the phototropic response.

**Note:**

Link to Learning



Use the navigation menu in the left panel of this [website](#) to view images of plants in motion.

## Plant Responses to Gravity

Whether or not they germinate in the light or in total darkness, shoots usually sprout up from the ground, and roots grow downward into the ground. A plant laid on its side in the dark will send shoots upward when given enough time. Gravitropism ensures that roots grow into the soil and that shoots grow toward sunlight. Growth of the shoot apical tip upward is called **negative gravitropism**, whereas growth of the roots downward is called **positive gravitropism**.

**Amyloplasts** (also known as **statoliths**) are specialized plastids that contain starch granules and settle downward in response to gravity. Amyloplasts are found in shoots and in specialized cells of the root cap. When a plant is tilted, the statoliths drop to the new bottom cell wall. A few hours later, the shoot or root will show growth in the new vertical direction.

The mechanism that mediates gravitropism is reasonably well understood. When amyloplasts settle to the bottom of the gravity-sensing cells in the



root or shoot, they physically contact the endoplasmic reticulum (ER), causing the release of calcium ions from inside the ER. This calcium signaling in the cells causes polar transport of the plant hormone IAA to the bottom of the cell. In roots, a high concentration of IAA inhibits cell elongation. The effect slows growth on the lower side of the root, while cells develop normally on the upper side. IAA has the opposite effect in shoots, where a higher concentration at the lower side of the shoot stimulates cell expansion, causing the shoot to grow up. After the shoot or root begin to grow vertically, the amyloplasts return to their normal position. Other hypotheses—involving the entire cell in the gravitropism effect—have been proposed to explain why some mutants that lack amyloplasts may still exhibit a weak gravitropic response.

## Growth Responses

A plant's sensory response to external stimuli relies on chemical messengers (hormones). Plant hormones affect all aspects of plant life, from flowering to fruit setting and maturation, and from phototropism to leaf fall. Potentially every cell in a plant can produce plant hormones. They can act in their cell of origin or be transported to other portions of the plant body, with many plant responses involving the synergistic or antagonistic interaction of two or more hormones. In contrast, animal hormones are produced in specific glands and transported to a distant site for action, and they act alone.

Plant hormones are a group of unrelated chemical substances that affect plant morphogenesis. Five major plant hormones are traditionally described: auxins (particularly IAA), cytokinins, gibberellins, ethylene, and abscisic acid. In addition, other nutrients and environmental conditions can be characterized as growth factors.

## Auxins

The term auxin is derived from the Greek word *auxein*, which means "to grow." **Auxins** are the main hormones responsible for cell elongation in

phototropism and gravitropism. They also control the differentiation of meristem into vascular tissue, and promote leaf development and arrangement. While many synthetic auxins are used as herbicides, IAA is the only naturally occurring auxin that shows physiological activity. Apical dominance—the inhibition of lateral bud formation—is triggered by auxins produced in the apical meristem. Flowering, fruit setting and ripening, and inhibition of **abscission** (leaf falling) are other plant responses under the direct or indirect control of auxins. Auxins also act as a relay for the effects of the blue light and red/far-red responses.

Commercial use of auxins is widespread in plant nurseries and for crop production. IAA is used as a rooting hormone to promote growth of adventitious roots on cuttings and detached leaves. Applying synthetic auxins to tomato plants in greenhouses promotes normal fruit development. Outdoor application of auxin promotes synchronization of fruit setting and dropping to coordinate the harvesting season. Fruits such as seedless cucumbers can be induced to set fruit by treating unfertilized plant flowers with auxins.

## Cytokinins

The effect of cytokinins was first reported when it was found that adding the liquid endosperm of coconuts to developing plant embryos in culture stimulated their growth. The stimulating growth factor was found to be **cytokinin**, a hormone that promotes cytokinesis (cell division). Almost 200 naturally occurring or synthetic cytokinins are known to date. Cytokinins are most abundant in growing tissues, such as roots, embryos, and fruits, where cell division is occurring. Cytokinins are known to delay senescence in leaf tissues, promote mitosis, and stimulate differentiation of the meristem in shoots and roots. Many effects on plant development are under the influence of cytokinins, either in conjunction with auxin or another hormone. For example, apical dominance seems to result from a balance between auxins that inhibit lateral buds, and cytokinins that promote bushier growth.

## **Gibberellins**

**Gibberellins** (GAs) are a group of about 125 closely related plant hormones that stimulate shoot elongation, seed germination, and fruit and flower maturation. GAs are synthesized in the root and stem apical meristems, young leaves, and seed embryos. In urban areas, GA antagonists are sometimes applied to trees under power lines to control growth and reduce the frequency of pruning.

GAs break dormancy (a state of inhibited growth and development) in the seeds of plants that require exposure to cold or light to germinate. Absciscic acid is a strong antagonist of GA action. Other effects of GAs include gender expression, seedless fruit development, and the delay of senescence in leaves and fruit. Seedless grapes are obtained through standard breeding methods and contain inconspicuous seeds that fail to develop. Because GAs are produced by the seeds, and because fruit development and stem elongation are under GA control, these varieties of grapes would normally produce small fruit in compact clusters. Maturing grapes are routinely treated with GA to promote larger fruit size, as well as looser bunches (longer stems), which reduces the instance of mildew infection ([\[link\]](#)).



In grapes, application of gibberellic acid increases the size of fruit and loosens clustering.  
(credit: Bob Nichols, USDA)

## **Abscisic Acid**

The plant hormone **abscisic acid** (ABA) was first discovered as the agent that causes the abscission or dropping of cotton bolls. However, more recent studies indicate that ABA plays only a minor role in the abscission process. ABA accumulates as a response to stressful environmental conditions, such as dehydration, cold temperatures, or shortened day lengths. Its activity counters many of the growth-promoting effects of GAs and auxins. ABA inhibits stem elongation and induces dormancy in lateral buds.

ABA induces dormancy in seeds by blocking germination and promoting the synthesis of storage proteins. Plants adapted to temperate climates require a long period of cold temperature before seeds germinate. This mechanism protects young plants from sprouting too early during unseasonably warm weather in winter. As the hormone gradually breaks down over winter, the seed is released from dormancy and germinates when conditions are favorable in spring. Another effect of ABA is to promote the development of winter buds; it mediates the conversion of the apical meristem into a dormant bud. Low soil moisture causes an increase in ABA, which causes stomata to close, reducing water loss in winter buds.

## **Ethylene**

**Ethylene** is associated with fruit ripening, flower wilting, and leaf fall. Ethylene is unusual because it is a volatile gas (C<sub>2</sub>H<sub>4</sub>). Hundreds of years ago, when gas street lamps were installed in city streets, trees that grew close to lamp posts developed twisted, thickened trunks and shed their leaves earlier than expected. These effects were caused by ethylene volatilizing from the lamps.

Aging tissues (especially senescing leaves) and nodes of stems produce ethylene. The best-known effect of the hormone, however, is the promotion of fruit ripening. Ethylene stimulates the conversion of starch and acids to sugars. Some people store unripe fruit, such as avocados, in a sealed paper bag to accelerate ripening; the gas released by the first fruit to mature will speed up the maturation of the remaining fruit. Ethylene also triggers leaf and fruit abscission, flower fading and dropping, and promotes germination in some cereals and sprouting of bulbs and potatoes.

Ethylene is widely used in agriculture. Commercial fruit growers control the timing of fruit ripening with application of the gas. Horticulturalists inhibit leaf dropping in ornamental plants by removing ethylene from greenhouses using fans and ventilation.

## Nontraditional Hormones

Recent research has discovered a number of compounds that also influence plant development. Their roles are less understood than the effects of the major hormones described so far.

**Jasmonates** play a major role in defense responses to herbivory. Their levels increase when a plant is wounded by a predator, resulting in an increase in toxic secondary metabolites. They contribute to the production of volatile compounds that attract natural enemies of predators. For example, chewing of tomato plants by caterpillars leads to an increase in jasmonic acid levels, which in turn triggers the release of volatile compounds that attract predators of the pest.

**Oligosaccharins** also play a role in plant defense against bacterial and fungal infections. They act locally at the site of injury, and can also be transported to other tissues. **Strigolactones** promote seed germination in some species and inhibit lateral apical development in the absence of auxins. Strigolactones also play a role in the establishment of mycorrhizae, a mutualistic association of plant roots and fungi. Brassinosteroids are important to many developmental and physiological processes. Signals between these compounds and other hormones, notably auxin and GAs, amplifies their physiological effect. Apical dominance, seed germination, gravitropism, and resistance to freezing are all positively influenced by hormones. Root growth and fruit dropping are inhibited by steroids.

## Plant Responses to Wind and Touch

The shoot of a pea plant winds around a trellis, while a tree grows on an angle in response to strong prevailing winds. These are examples of how plants respond to touch or wind.

The movement of a plant subjected to constant directional pressure is called **thigmotropism**, from the Greek words *thigma* meaning “touch,” and *tropism* implying “direction.” Tendrils are one example of this. The meristematic region of tendrils is very touch sensitive; light touch will evoke a quick coiling response. Cells in contact with a support surface

contract, whereas cells on the opposite side of the support expand ([\[link\]](#)). Application of jasmonic acid is sufficient to trigger tendril coiling without a mechanical stimulus.

A **thigmonastic** response is a touch response independent of the direction of stimulus [\[link\]](#). In the Venus flytrap, two modified leaves are joined at a hinge and lined with thin fork-like tines along the outer edges. Tiny hairs are located inside the trap. When an insect brushes against these trigger hairs, touching two or more of them in succession, the leaves close quickly, trapping the prey. Glands on the leaf surface secrete enzymes that slowly digest the insect. The released nutrients are absorbed by the leaves, which reopen for the next meal.

**Thigmomorphogenesis** is a slow developmental change in the shape of a plant subjected to continuous mechanical stress. When trees bend in the wind, for example, growth is usually stunted and the trunk thickens. Strengthening tissue, especially xylem, is produced to add stiffness to resist the wind's force. Researchers hypothesize that mechanical strain induces growth and differentiation to strengthen the tissues. Ethylene and jasmonate are likely involved in thigmomorphogenesis.

**Note:**

Link to Learning



Use the menu at the left to navigate to three short [movies](#): a Venus fly trap capturing prey, the progressive closing of sensitive plant leaflets, and the twining of tendrils.

## **Defense Responses against Herbivores and Pathogens**

Plants face two types of enemies: herbivores and pathogens. Herbivores both large and small use plants as food, and actively chew them. Pathogens are agents of disease. These infectious microorganisms, such as fungi, bacteria, and nematodes, live off of the plant and damage its tissues. Plants have developed a variety of strategies to discourage or kill attackers.

The first line of defense in plants is an intact and impenetrable barrier. Bark and the waxy cuticle can protect against predators. Other adaptations against herbivory include thorns, which are modified branches, and spines, which are modified leaves. They discourage animals by causing physical damage and inducing rashes and allergic reactions. A plant's exterior protection can be compromised by mechanical damage, which may provide an entry point for pathogens. If the first line of defense is breached, the plant must resort to a different set of defense mechanisms, such as toxins and enzymes.

Secondary metabolites are compounds that are not directly derived from photosynthesis and are not necessary for respiration or plant growth and development. Many metabolites are toxic, and can even be lethal to animals that ingest them. Some metabolites are alkaloids, which discourage predators with noxious odors (such as the volatile oils of mint and sage) or repellent tastes (like the bitterness of quinine). Other alkaloids affect herbivores by causing either excessive stimulation (caffeine is one example) or the lethargy associated with opioids. Some compounds become toxic after ingestion; for instance, glycol cyanide in the cassava root releases cyanide only upon ingestion by the herbivore.

Mechanical wounding and predator attacks activate defense and protection mechanisms both in the damaged tissue and at sites farther from the injury location. Some defense reactions occur within minutes: others over several hours. The infected and surrounding cells may die, thereby stopping the spread of infection.

Long-distance signaling elicits a systemic response aimed at deterring the predator. As tissue is damaged, jasmonates may promote the synthesis of compounds that are toxic to predators. Jasmonates also elicit the synthesis



of volatile compounds that attract parasitoids, which are insects that spend their developing stages in or on another insect, and eventually kill their host. The plant may activate abscission of injured tissue if it is damaged beyond repair.

## **Section Summary**

Plants respond to light by changes in morphology and activity. Irradiation by red light converts the photoreceptor phytochrome to its far-red light-absorbing form—Pfr. This form controls germination and flowering in response to length of day, as well as triggers photosynthesis in dormant plants or those that just emerged from the soil. Blue-light receptors, cryptochromes, and phototropins are responsible for phototropism. Amyloplasts, which contain heavy starch granules, sense gravity. Shoots exhibit negative gravitropism, whereas roots exhibit positive gravitropism. Plant hormones—naturally occurring compounds synthesized in small amounts—can act both in the cells that produce them and in distant tissues and organs. Auxins are responsible for apical dominance, root growth, directional growth toward light, and many other growth responses. Cytokinins stimulate cell division and counter apical dominance in shoots. Gibberellins inhibit dormancy of seeds and promote stem growth. Absciscic acid induces dormancy in seeds and buds, and protects plants from excessive water loss by promoting stomatal closure. Ethylene gas speeds up fruit ripening and dropping of leaves. Plants respond to touch by rapid movements (thigmotropy and thigmonasty) and slow differential growth (thigmomorphogenesis). Plants have evolved defense mechanisms against predators and pathogens. Physical barriers like bark and spines protect tender tissues. Plants also have chemical defenses, including toxic secondary metabolites and hormones, which elicit additional defense mechanisms.

## **Review Questions**

### **Exercise:**

**Problem:**

The main photoreceptor that triggers phototropism is a \_\_\_\_\_.

- a. phytochrome
- b. cryptochrome
- c. phototropin
- d. carotenoid

---

**Solution:**

C

**Exercise:**

**Problem:** Phytochrome is a plant pigment protein that:

- a. mediates plant infection
- b. promotes plant growth
- c. mediates morphological changes in response to red and far-red light
- d. inhibits plant growth

---

**Solution:**

C

**Exercise:****Problem:**

A mutant plant has roots that grow in all directions. Which of the following organelles would you expect to be missing in the cell?

- a. mitochondria
- b. amyloplast
- c. chloroplast

d. nucleus

---

**Solution:**

B

**Exercise:**

**Problem:**

After buying green bananas or unripe avocados, they can be kept in a brown bag to ripen. The hormone released by the fruit and trapped in the bag is probably:

- a. abscisic acid
- b. cytokinin
- c. ethylene
- d. gibberellic acid

---

**Solution:**

C

**Exercise:**

**Problem:**

A decrease in the level of which hormone releases seeds from dormancy?

- a. abscisic acid
- b. cytokinin
- c. ethylene
- d. gibberellic acid

---

**Solution:**

A

**Exercise:****Problem:**

A seedling germinating under a stone grows at an angle away from the stone and upward. This response to touch is called \_\_\_\_\_.

- a. gravitropism
- b. thigmonasty
- c. thigmotropism
- d. skototropism

---

**Solution:**

C

**Free Response****Exercise:****Problem:**

Owners and managers of plant nurseries have to plan lighting schedules for a long-day plant that will flower in February. What lighting periods will be most effective? What color of light should be chosen?

---

**Solution:**

A long-day plant needs a higher proportion of the Pfr form to Pr form of phytochrome. The plant requires long periods of illumination with light enriched in the red range of the spectrum.

**Exercise:**

**Problem:**

What are the major benefits of gravitropism for a germinating seedling?

---

**Solution:**

Gravitropism will allow roots to dig deep into the soil to find water and minerals, whereas the seedling will grow towards light to enable photosynthesis.

**Exercise:****Problem:**

Fruit and vegetable storage facilities are usually refrigerated and well ventilated. Why are these conditions advantageous?

---

**Solution:**

Refrigeration slows chemical reactions, including fruit maturation. Ventilation removes the ethylene gas that speeds up fruit ripening.

**Exercise:****Problem:**

Stomata close in response to bacterial infection. Why is this response a mechanism of defense for the plant? Which hormone is most likely to mediate this response?

---

**Solution:**

To prevent further entry of pathogens, stomata close, even if they restrict entry of CO<sub>2</sub>. Some pathogens secrete virulence factors that inhibit the closing of stomata. Abscissic acid is the stress hormone responsible for inducing closing of stomata.

**Glossary**

abscisic acid (ABA)

plant hormone that induces dormancy in seeds and other organs

abscission

physiological process that leads to the fall of a plant organ (such as leaf or petal drop)

auxin

plant hormone that influences cell elongation (in phototropism), gravitropism, apical dominance and root growth

chromophore

molecule that absorbs light

cryptochrome

protein that absorbs light in the blue and ultraviolet regions of the light spectrum

cytokinin

plant hormone that promotes cell division

ethylene

volatile plant hormone that is associated with fruit ripening, flower wilting, and leaf fall

gibberellin (GA)

plant hormone that stimulates shoot elongation, seed germination, and the maturation and dropping of fruit and flowers

jasmonates

small family of compounds derived from the fatty acid linoleic acid

negative gravitropism

growth away from Earth's gravity

oligosaccharin

hormone important in plant defenses against bacterial and fungal infections

photomorphogenesis

growth and development of plants in response to light

photoperiodism

occurrence of plant processes, such as germination and flowering, according to the time of year

phototropin

blue-light receptor that promotes phototropism, stomatal opening and closing, and other responses that promote photosynthesis

phototropism

directional bending of a plant toward a light source

phytochrome

plant pigment protein that exists in two reversible forms (Pr and Pfr) and mediates morphologic changes in response to red light

positive gravitropism

growth toward Earth's gravitational center

statolith

(also, **amyloplast**) plant organelle that contains heavy starch granules

strigolactone

hormone that promotes seed germination in some species and inhibits lateral apical development in the absence of auxins

thigmomorphogenesis

developmental response to touch

thigmonastic

directional growth of a plant independent of the direction in which contact is applied

thigmotropism

directional growth of a plant in response to constant contact

## Introduction

class="introduction"

For this (a) squash seedling (*Cucurbita maxima*) to develop into a mature plant bearing its (b) fruit, numerous nutritional requirements must be met. (credit a: modification of work by Julian Colton; credit b: modification of work by "Wildfeuer"/Wikimedia Commons)



(a)



(b)

Cucurbitaceae is a family of plants first cultivated in Mesoamerica, although several species are native to North America. The family includes many edible species, such as squash and pumpkin, as well as inedible gourds. In order to grow and develop into mature, fruit-bearing plants, many requirements must be met and events must be coordinated. Seeds must germinate under the right conditions in the soil; therefore, temperature, moisture, and soil quality are important factors that play a role



in germination and seedling development. Soil quality and climate are significant to plant distribution and growth. The young seedling will eventually grow into a mature plant, and the roots will absorb nutrients and water from the soil. At the same time, the aboveground parts of the plant will absorb carbon dioxide from the atmosphere and use energy from sunlight to produce organic compounds through photosynthesis. This chapter will explore the complex dynamics between plants and soils, and the adaptations that plants have evolved to make better use of nutritional resources.

## Nutritional Requirements of Plants

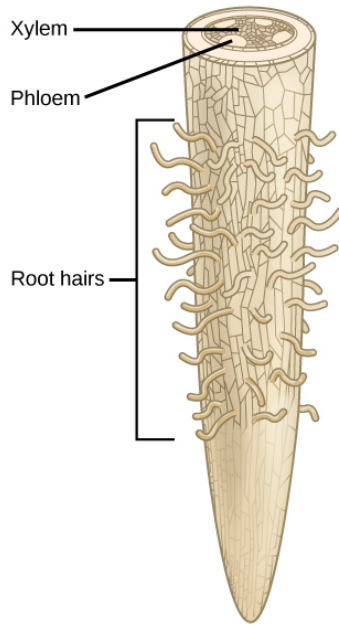
By the end of this section, you will be able to:

- Describe how plants obtain nutrients
- List the elements and compounds required for proper plant nutrition
- Describe an essential nutrient

Plants are unique organisms that can absorb nutrients and water through their root system, as well as carbon dioxide from the atmosphere. Soil quality and climate are the major determinants of plant distribution and growth. The combination of soil nutrients, water, and carbon dioxide, along with sunlight, allows plants to grow.

## The Chemical Composition of Plants

Since plants require nutrients in the form of elements such as carbon and potassium, it is important to understand the chemical composition of plants. The majority of volume in a plant cell is water; it typically comprises 80 to 90 percent of the plant's total weight. Soil is the water source for land plants, and can be an abundant source of water, even if it appears dry. Plant roots absorb water from the soil through root hairs and transport it up to the leaves through the xylem. As water vapor is lost from the leaves, the process of transpiration and the polarity of water molecules (which enables them to form hydrogen bonds) draws more water from the roots up through the plant to the leaves ([link](#)). Plants need water to support cell structure, for metabolic functions, to carry nutrients, and for photosynthesis.



Water is absorbed through the root hairs and moves up the xylem to the leaves.

Plant cells need essential substances, collectively called nutrients, to sustain life. Plant nutrients may be composed of either organic or inorganic compounds. An **organic compound** is a chemical compound that contains carbon, such as carbon dioxide obtained from the atmosphere. Carbon that was obtained from atmospheric CO<sub>2</sub> composes the majority of the dry mass within most plants. An **inorganic compound** does not contain carbon and is not part of, or produced by, a living organism. Inorganic substances, which form the majority of the soil solution, are commonly called minerals: those required by plants include nitrogen (N) and potassium (K) for structure and regulation.

## Essential Nutrients

Plants require only light, water and about 20 elements to support all their biochemical needs: these 20 elements are called essential nutrients ([link](#)). For an element to be regarded as **essential**, three criteria are required: 1) a

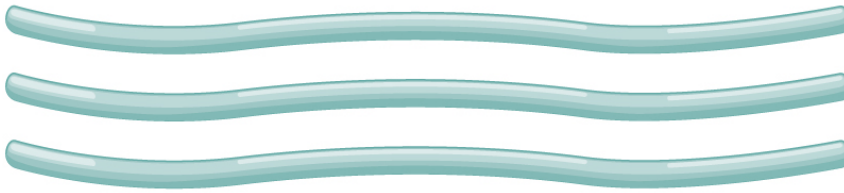
plant cannot complete its life cycle without the element; 2) no other element can perform the function of the element; and 3) the element is directly involved in plant nutrition.

Essential Elements for Plant Growth	
Macronutrients	Micronutrients
Carbon (C)	Iron (Fe)
Hydrogen (H)	Manganese (Mn)
Oxygen (O)	Boron (B)
Nitrogen (N)	Molybdenum (Mo)
Phosphorus (P)	Copper (Cu)
Potassium (K)	Zinc (Zn)
Calcium (Ca)	Chlorine (Cl)
Magnesium (Mg)	Nickel (Ni)
Sulfur (S)	Cobalt (Co)
	Sodium (Na)
	Silicon (Si)

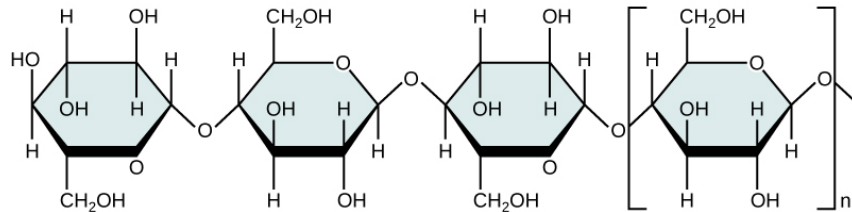
## Macronutrients and Micronutrients

The essential elements can be divided into two groups: macronutrients and micronutrients. Nutrients that plants require in larger amounts are called **macronutrients**. About half of the essential elements are considered macronutrients: carbon, hydrogen, oxygen, nitrogen, phosphorus, potassium, calcium, magnesium and sulfur. The first of these macronutrients, carbon (C), is required to form carbohydrates, proteins, nucleic acids, and many other compounds; it is therefore present in all macromolecules. On average, the dry weight (excluding water) of a cell is 50 percent carbon. As shown in [\[link\]](#), carbon is a key part of plant biomolecules.

Cellulose fibers



Cellulose structure



Cellulose, the main structural component of the plant cell wall, makes up over thirty percent of plant matter. It is the most abundant organic compound on earth.

The next most abundant element in plant cells is nitrogen (N); it is part of proteins and nucleic acids. Nitrogen is also used in the synthesis of some vitamins. Hydrogen and oxygen are macronutrients that are part of many

organic compounds, and also form water. Oxygen is necessary for cellular respiration; plants use oxygen to store energy in the form of ATP.

Phosphorus (P), another macromolecule, is necessary to synthesize nucleic acids and phospholipids. As part of ATP, phosphorus enables food energy to be converted into chemical energy through oxidative phosphorylation.

Likewise, light energy is converted into chemical energy during photophosphorylation in photosynthesis, and into chemical energy to be extracted during respiration. Sulfur is part of certain amino acids, such as cysteine and methionine, and is present in several coenzymes. Sulfur also plays a role in photosynthesis as part of the electron transport chain, where hydrogen gradients play a key role in the conversion of light energy into ATP. Potassium (K) is important because of its role in regulating stomatal opening and closing. As the openings for gas exchange, stomata help maintain a healthy water balance; a potassium ion pump supports this process.

Magnesium (Mg) and calcium (Ca) are also important macronutrients. The role of calcium is twofold: to regulate nutrient transport, and to support many enzyme functions. Magnesium is important to the photosynthetic process. These minerals, along with the micronutrients, which are described below, also contribute to the plant's ionic balance.

In addition to macronutrients, organisms require various elements in small amounts. These **micronutrients**, or trace elements, are present in very small quantities. They include boron (B), chlorine (Cl), manganese (Mn), iron (Fe), zinc (Zn), copper (Cu), molybdenum (Mo), nickel (Ni), silicon (Si), and sodium (Na).

Deficiencies in any of these nutrients—particularly the macronutrients—can adversely affect plant growth ([link](#)). Depending on the specific nutrient, a lack can cause stunted growth, slow growth, or chlorosis (yellowing of the leaves). Extreme deficiencies may result in leaves showing signs of cell death.

**Note:**

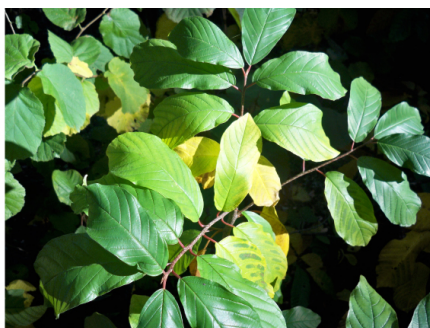
[Link to Learning](#)



Visit this [website](#) to participate in an interactive experiment on plant nutrient deficiencies. You can adjust the amounts of N, P, K, Ca, Mg, and Fe that plants receive . . . and see what happens.



(a)



(b)



(c)



(d)

Nutrient deficiency is evident in the symptoms these plants show. This (a) grape tomato suffers from blossom end rot caused by calcium deficiency. The yellowing in this (b) *Frangula alnus* results from magnesium deficiency.

Inadequate magnesium also leads to (c) interveinal chlorosis, seen here in a sweetgum

leaf. This (d) palm is affected by potassium deficiency. (credit c: modification of work by Jim Conrad; credit d: modification of work by Malcolm Manners)

**Note:**

**Everyday Connection**

**Hydroponics**

Hydroponics is a method of growing plants in a water-nutrient solution instead of soil. Since its advent, hydroponics has developed into a growing process that researchers often use. Scientists who are interested in studying plant nutrient deficiencies can use hydroponics to study the effects of different nutrient combinations under strictly controlled conditions.

Hydroponics has also developed as a way to grow flowers, vegetables, and other crops in greenhouse environments. You might find hydroponically grown produce at your local grocery store. Today, many lettuces and tomatoes in your market have been hydroponically grown.

## **Section Summary**

Plants can absorb inorganic nutrients and water through their root system, and carbon dioxide from the environment. The combination of organic compounds, along with water, carbon dioxide, and sunlight, produce the energy that allows plants to grow. Inorganic compounds form the majority of the soil solution. Plants access water through the soil. Water is absorbed by the plant root, transports nutrients throughout the plant, and maintains the structure of the plant. Essential elements are indispensable elements for plant growth. They are divided into macronutrients and micronutrients. The macronutrients plants require are carbon, nitrogen, hydrogen, oxygen, phosphorus, potassium, calcium, magnesium, and sulfur. Important



micronutrients include iron, manganese, boron, molybdenum, copper, zinc, chlorine, nickel, cobalt, silicon and sodium.

## Review Questions

### Exercise:

#### Problem:

For an element to be regarded as essential, all of the following criteria must be met, except:

- a. No other element can perform the function.
- b. The element is directly involved in plant nutrition.
- c. The element is inorganic.
- d. The plant cannot complete its lifecycle without the element.

---

#### Solution:

C

### Exercise:

#### Problem:

The nutrient that is part of carbohydrates, proteins, and nucleic acids, and that forms biomolecules, is \_\_\_\_\_.

- a. nitrogen
- b. carbon
- c. magnesium
- d. iron

---

#### Solution:

B

### Exercise:

**Problem:** Most \_\_\_\_\_ are necessary for enzyme function.

- a. micronutrients
- b. macronutrients
- c. biomolecules
- d. essential nutrients

---

**Solution:**

A

**Exercise:**

**Problem:** What is the main water source for land plants?

- a. rain
- b. soil
- c. biomolecules
- d. essential nutrients

---

**Solution:**

B

**Free Response**

**Exercise:**

**Problem:**

What type of plant problems result from nitrogen and calcium deficiencies?

---

**Solution:**

Deficiencies in these nutrients could result in stunted growth, slow growth, and chlorosis.

**Exercise:**

**Problem:**

Research the life of Jan Babtista van Helmont. What did the van Helmont experiment show?

---

**Solution:**

van Helmont showed that plants do not consume soil, which is correct. He also thought that plant growth and increased weight resulted from the intake of water, a conclusion that has since been disproven.

**Exercise:**

**Problem:**

List two essential macronutrients and two essential nutrients.

---

**Solution:**

Answers may vary. Essential macronutrients include carbon, hydrogen, oxygen, nitrogen, phosphorus, potassium, calcium, magnesium, and sulfur. Essential micronutrients include iron, manganese, boron, molybdenum, copper, zinc, chlorine, nickel, cobalt, sodium, and silicon.

## **Glossary**

inorganic compound

chemical compound that does not contain carbon; it is not part of or produced by a living organism

macronutrient

nutrient that is required in large amounts for plant growth; carbon, hydrogen, oxygen, nitrogen, phosphorus, potassium, calcium,

magnesium, and sulfur

micronutrient

nutrient required in small amounts; also called trace element

organic compound

chemical compound that contains carbon

## The Soil

By the end of this section, you will be able to:

- Describe how soils are formed
- Explain soil composition
- Describe a soil profile

Plants obtain inorganic elements from the soil, which serves as a natural medium for land plants. **Soil** is the outer loose layer that covers the surface of Earth. Soil quality is a major determinant, along with climate, of plant distribution and growth. Soil quality depends not only on the chemical composition of the soil, but also the topography (regional surface features) and the presence of living organisms. In agriculture, the history of the soil, such as the cultivating practices and previous crops, modify the characteristics and fertility of that soil.

Soil develops very slowly over long periods of time, and its formation results from natural and environmental forces acting on mineral, rock, and organic compounds. Soils can be divided into two groups: **organic soils** are those that are formed from sedimentation and primarily composed of organic matter, while those that are formed from the weathering of rocks and are primarily composed of inorganic material are called **mineral soils**. Mineral soils are predominant in terrestrial ecosystems, where soils may be covered by water for part of the year or exposed to the atmosphere.

## Soil Composition

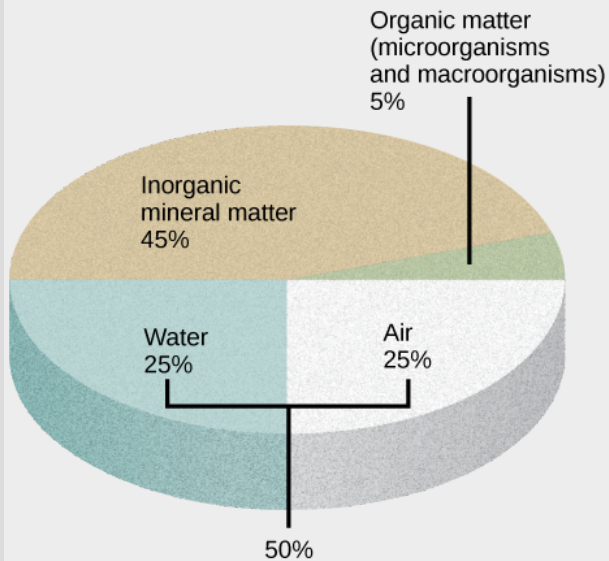
Soil consists of these major components ([link](#)):

- inorganic mineral matter, about 40 to 45 percent of the soil volume
- organic matter, about 5 percent of the soil volume
- water and air, about 50 percent of the soil volume

The amount of each of the four major components of soil depends on the amount of vegetation, soil compaction, and water present in the soil. A good healthy soil has sufficient air, water, minerals, and organic material to promote and sustain plant life.

**Note:**

**Art Connection**



The four major components of soil are shown: inorganic minerals, organic matter, water, and air.

Soil compaction can result when soil is compressed by heavy machinery or even foot traffic. How might this compaction change the soil composition?

The organic material of soil, called **humus**, is made up of microorganisms (dead and alive), and dead animals and plants in varying stages of decay. Humus improves soil structure and provides plants with water and minerals. The inorganic material of soil consists of rock, slowly broken down into smaller particles that vary in size. Soil particles that are 0.1 to 2 mm in diameter are **sand**. Soil particles between 0.002 and 0.1 mm are called **silt**, and even smaller particles, less than 0.002 mm in diameter, are called **clay**. Some soils have no dominant particle size and contain a mixture of sand, silt, and humus; these soils are called **loams**.

**Note:****Link to Learning**

Explore this [interactive map](#) from the USDA's National Cooperative Soil Survey to access soil data for almost any region in the United States.

## Soil Formation

Soil formation is the consequence of a combination of biological, physical, and chemical processes. Soil should ideally contain 50 percent solid material and 50 percent pore space. About one-half of the pore space should contain water, and the other half should contain air. The organic component of soil serves as a cementing agent, returns nutrients to the plant, allows soil to store moisture, makes soil tillable for farming, and provides energy for soil microorganisms. Most soil microorganisms—bacteria, algae, or fungi—are dormant in dry soil, but become active once moisture is available.

Soil distribution is not homogenous because its formation results in the production of layers; together, the vertical section of a soil is called the **soil profile**. Within the soil profile, soil scientists define zones called horizons. A **horizon** is a soil layer with distinct physical and chemical properties that differ from those of other layers. Five factors account for soil formation: parent material, climate, topography, biological factors, and time.

### Parent Material

The organic and inorganic material in which soils form is the **parent material**. Mineral soils form directly from the weathering of **bedrock**, the

solid rock that lies beneath the soil, and therefore, they have a similar composition to the original rock. Other soils form in materials that came from elsewhere, such as sand and glacial drift. Materials located in the depth of the soil are relatively unchanged compared with the deposited material. Sediments in rivers may have different characteristics, depending on whether the stream moves quickly or slowly. A fast-moving river could have sediments of rocks and sand, whereas a slow-moving river could have fine-textured material, such as clay.

## **Climate**

Temperature, moisture, and wind cause different patterns of weathering and therefore affect soil characteristics. The presence of moisture and nutrients from weathering will also promote biological activity: a key component of a quality soil.

## **Topography**

Regional surface features (familiarily called “the lay of the land”) can have a major influence on the characteristics and fertility of a soil. Topography affects water runoff, which strips away parent material and affects plant growth. Steeps soils are more prone to erosion and may be thinner than soils that are relatively flat or level.

## **Biological factors**

The presence of living organisms greatly affects soil formation and structure. Animals and microorganisms can produce pores and crevices, and plant roots can penetrate into crevices to produce more fragmentation. Plant secretions promote the development of microorganisms around the root, in an area known as the **rhizosphere**. Additionally, leaves and other material that fall from plants decompose and contribute to soil composition.



## Time

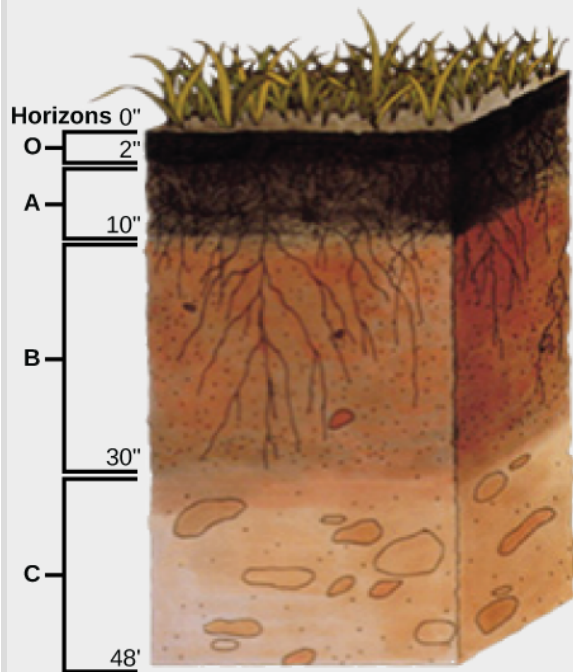
Time is an important factor in soil formation because soils develop over long periods. Soil formation is a dynamic process. Materials are deposited over time, decompose, and transform into other materials that can be used by living organisms or deposited onto the surface of the soil.

## Physical Properties of the Soil

Soils are named and classified based on their horizons. The soil profile has four distinct layers: 1) O horizon; 2) A horizon; 3) B horizon, or subsoil; and 4) C horizon, or soil base ([\[link\]](#)). The **O horizon** has freshly decomposing organic matter—humus—at its surface, with decomposed vegetation at its base. Humus enriches the soil with nutrients and enhances soil moisture retention. Topsoil—the top layer of soil—is usually two to three inches deep, but this depth can vary considerably. For instance, river deltas like the Mississippi River delta have deep layers of topsoil. Topsoil is rich in organic material; microbial processes occur there, and it is the “workhorse” of plant production. The **A horizon** consists of a mixture of organic material with inorganic products of weathering, and it is therefore the beginning of true mineral soil. This horizon is typically darkly colored because of the presence of organic matter. In this area, rainwater percolates through the soil and carries materials from the surface. The **B horizon** is an accumulation of mostly fine material that has moved downward, resulting in a dense layer in the soil. In some soils, the B horizon contains nodules or a layer of calcium carbonate. The **C horizon**, or soil base, includes the parent material, plus the organic and inorganic material that is broken down to form soil. The parent material may be either created in its natural place, or transported from elsewhere to its present location. Beneath the C horizon lies bedrock.

### Note:

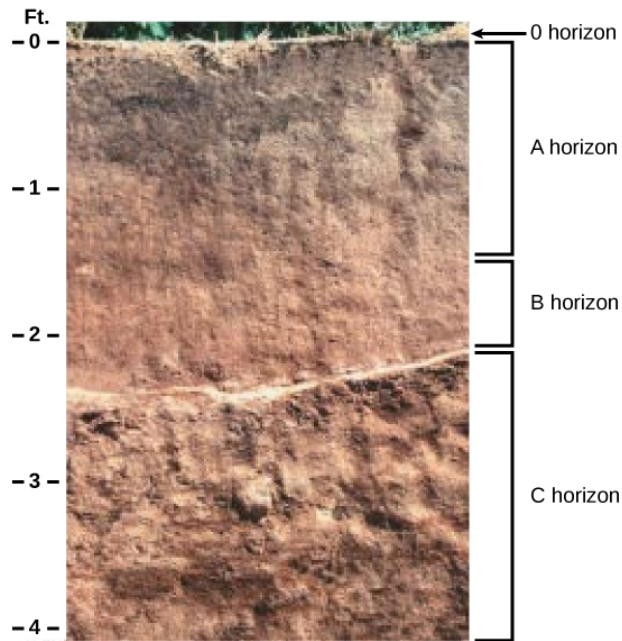
Art Connection



This soil profile shows the different soil layers (O horizon, A horizon, B horizon, and C horizon) found in typical soils. (credit: modification of work by USDA)

Which horizon is considered the topsoil, and which is considered the subsoil?

Some soils may have additional layers, or lack one of these layers. The thickness of the layers is also variable, and depends on the factors that influence soil formation. In general, immature soils may have O, A, and C horizons, whereas mature soils may display all of these, plus additional layers ([link](#)).



The San Joaquin soil profile has an O horizon, A horizon, B horizon, and C horizon. (credit: modification of work by USDA)

**Note:**

**Career Connections**

**Soil Scientist**

A soil scientist studies the biological components, physical and chemical properties, distribution, formation, and morphology of soils. Soil scientists need to have a strong background in physical and life sciences, plus a foundation in mathematics. They may work for federal or state agencies, academia, or the private sector. Their work may involve collecting data, carrying out research, interpreting results, inspecting soils, conducting soil surveys, and recommending soil management programs.



This soil scientist is studying the horizons and composition of soil at a research site. (credit: USDA)

Many soil scientists work both in an office and in the field. According to the United States Department of Agriculture (USDA): “a soil scientist needs good observation skills to analyze and determine the characteristics of different types of soils. Soil types are complex and the geographical areas a soil scientist may survey are varied. Aerial photos or various satellite images are often used to research the areas. Computer skills and geographic information systems (GIS) help the scientist to analyze the multiple facets of geomorphology, topography, vegetation, and climate to discover the patterns left on the landscape.”<sup>[footnote]</sup> Soil scientists play a key role in understanding the soil’s past, analyzing present conditions, and making recommendations for future soil-related practices.

National Resources Conservation Service / United States Department of Agriculture. “Careers in Soil Science.”

<http://soils.usda.gov/education/facts/careers.html>

## Section Summary

Plants obtain mineral nutrients from the soil. Soil is the outer loose layer that covers the surface of Earth. Soil quality depends on the chemical composition of the soil, the topography, the presence of living organisms, the climate, and time. Agricultural practice and history may also modify the characteristics and fertility of soil. Soil consists of four major components: 1) inorganic mineral matter, 2) organic matter, 3) water and air, and 4) living matter. The organic material of soil is made of humus, which improves soil structure and provides water and minerals. Soil inorganic material consists of rock slowly broken down into smaller particles that vary in size, such as sand, silt, and loam.

Soil formation results from a combination of biological, physical, and chemical processes. Soil is not homogenous because its formation results in the production of layers called a soil profile. Factors that affect soil formation include: parent material, climate, topography, biological factors, and time. Soils are classified based on their horizons, soil particle size, and proportions. Most soils have four distinct horizons: O, A, B, and C.

## Art Connections

### Exercise:

#### Problem:

[\[link\]](#) Soil compaction can result when soil is compressed by heavy machinery or even foot traffic. How might this compaction change the soil composition?

---

#### Solution:

[\[link\]](#) The air content of the soil decreases.

### Exercise:

**Problem:**

[\[link\]](#) Which horizon is considered the topsoil, and which is considered the subsoil?

---

**Solution:**

[\[link\]](#) The A horizon is the topsoil, and the B horizon is subsoil.

**Review Questions****Exercise:**

**Problem:** Which factors affect soil quality?

- a. chemical composition
  - b. history of the soil
  - c. presence of living organisms and topography
  - d. all of the above
- 

**Solution:**

D

**Exercise:****Problem:**

Soil particles that are 0.1 to 2 mm in diameter are called \_\_\_\_\_.

- a. sand
  - b. silt
  - c. clay
  - d. loam
- 

**Solution:**

A

**Exercise:**

**Problem:**

A soil consists of layers called \_\_\_\_\_ that taken together are called a \_\_\_\_\_.

- a. soil profiles : horizon
- b. horizons : soil profile
- c. horizons : humus
- d. humus : soil profile

---

**Solution:**

B

**Exercise:**

**Problem:**

What is the term used to describe the solid rock that lies beneath the soil?

- a. sand
- b. bedrock
- c. clay
- d. loam

---

**Solution:**

B

**Exercise:**

**Problem:**

Describe the main differences between a mineral soil and an organic soil.

---

**Solution:**

A mineral soil forms from the weathering of rocks; it is inorganic material. An organic soil is formed from sedimentation; it mostly consists of humus.

**Exercise:****Problem:**

Name and briefly explain the factors that affect soil formation.

---

**Solution:**

Parent material, climate, topography, biological factors, and time affect soil formation. Parent material is the material in which soils form. Climate describes how temperature, moisture, and wind cause different patterns of weathering, influencing the characteristics of the soil. Topography affects the characteristics and fertility of a soil. Biological factors include the presence of living organisms that greatly affect soil formation. Processes such as freezing and thawing may produce cracks in rocks; plant roots can penetrate these crevices and produce more fragmentation. Time affects soil because soil develops over long periods.

**Exercise:****Problem:**

Describe how topography influences the characteristics and fertility of a soil.

---

**Solution:**

Topography affects water runoff, which strips away parent material and affects plant growth. Steep soils are more prone to erosion and may be thinner than soils that are on level surfaces.



## **Glossary**

### **A horizon**

consists of a mixture of organic material with inorganic products of weathering

### **B horizon**

soil layer that is an accumulation of mostly fine material that has moved downward

### **bedrock**

solid rock that lies beneath the soil

### **C horizon**

layer of soil that contains the parent material, and the organic and inorganic material that is broken down to form soil; also known as the soil base

### **clay**

soil particles that are less than 0.002 mm in diameter

### **horizon**

soil layer with distinct physical and chemical properties, which differs from other layers depending on how and when it was formed

### **humus**

organic material of soil; made up of microorganisms, dead animals and plants in varying stages of decay

### **loam**

soil that has no dominant particle size

### **mineral soil**

type of soil that is formed from the weathering of rocks and inorganic material; composed primarily of sand, silt, and clay

### **O horizon**

layer of soil with humus at the surface and decomposed vegetation at the base

organic soil

type of soil that is formed from sedimentation; composed primarily of organic material

parent material

organic and inorganic material in which soils form

rhizosphere

area of soil affected by root secretions and microorganisms

sand

soil particles between 0.1–2 mm in diameter

silt

soil particles between 0.002 and 0.1 mm in diameter

soil profile

vertical section of a soil

soil

outer loose layer that covers the surface of Earth

## Nutritional Adaptations of Plants

By the end of this section, you will be able to:

- Understand the nutritional adaptations of plants
- Describe mycorrhizae
- Explain nitrogen fixation

Plants obtain food in two different ways. Autotrophic plants can make their own food from inorganic raw materials, such as carbon dioxide and water, through photosynthesis in the presence of sunlight. Green plants are included in this group. Some plants, however, are heterotrophic: they are totally parasitic and lacking in chlorophyll. These plants, referred to as holo-parasitic plants, are unable to synthesize organic carbon and draw all of their nutrients from the host plant.

Plants may also enlist the help of microbial partners in nutrient acquisition. Particular species of bacteria and fungi have evolved along with certain plants to create a mutualistic symbiotic relationship with roots. This improves the nutrition of both the plant and the microbe. The formation of nodules in legume plants and mycorrhization can be considered among the nutritional adaptations of plants. However, these are not the only type of adaptations that we may find; many plants have other adaptations that allow them to thrive under specific conditions.

### Note:

Link to Learning



This [video](#) reviews basic concepts about photosynthesis. In the left panel, click each tab to select a topic for review.

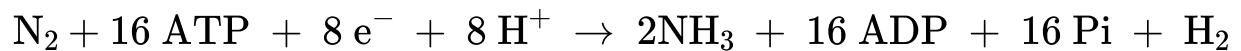
## Nitrogen Fixation: Root and Bacteria Interactions

Nitrogen is an important macronutrient because it is part of nucleic acids and proteins. Atmospheric nitrogen, which is the diatomic molecule  $N_2$ , or dinitrogen, is the largest pool of nitrogen in terrestrial ecosystems.

However, plants cannot take advantage of this nitrogen because they do not have the necessary enzymes to convert it into biologically useful forms.

However, nitrogen can be “fixed,” which means that it can be converted to ammonia ( $NH_3$ ) through biological, physical, or chemical processes. As you have learned, biological nitrogen fixation (BNF) is the conversion of atmospheric nitrogen ( $N_2$ ) into ammonia ( $NH_3$ ), exclusively carried out by prokaryotes such as soil bacteria or cyanobacteria. Biological processes contribute 65 percent of the nitrogen used in agriculture. The following equation represents the process:

**Equation:**



The most important source of BNF is the symbiotic interaction between soil bacteria and legume plants, including many crops important to humans ([link](#)). The  $NH_3$  resulting from fixation can be transported into plant tissue and incorporated into amino acids, which are then made into plant proteins. Some legume seeds, such as soybeans and peanuts, contain high levels of protein, and serve among the most important agricultural sources of protein in the world.

**Note:**

Art Connection



(a)

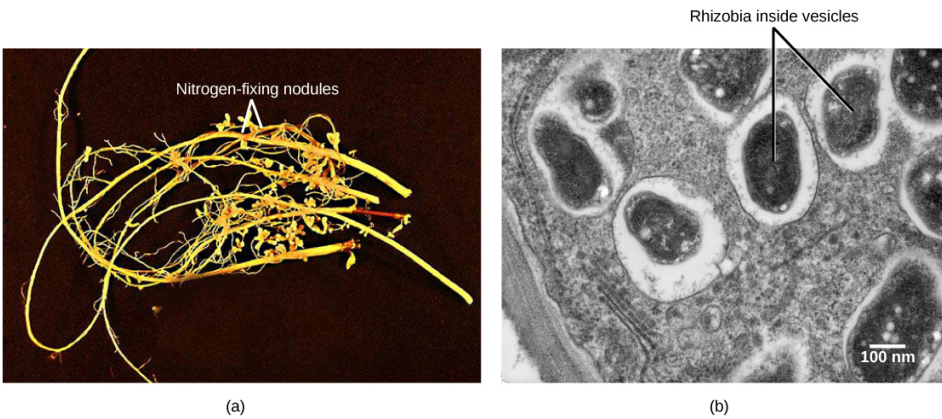
(b)

(c)

Some common edible legumes—like (a) peanuts, (b) beans, and (c) chickpeas—are able to interact symbiotically with soil bacteria that fix nitrogen. (credit a: modification of work by Jules Clancy; credit b: modification of work by USDA)

Farmers often rotate corn (a cereal crop) and soy beans (a legume), planting a field with each crop in alternate seasons. What advantage might this crop rotation confer?

Soil bacteria, collectively called **rhizobia**, symbiotically interact with legume roots to form specialized structures called **nodules**, in which nitrogen fixation takes place. This process entails the reduction of atmospheric nitrogen to ammonia, by means of the enzyme **nitrogenase**. Therefore, using rhizobia is a natural and environmentally friendly way to fertilize plants, as opposed to chemical fertilization that uses a nonrenewable resource, such as natural gas. Through symbiotic nitrogen fixation, the plant benefits from using an endless source of nitrogen from the atmosphere. The process simultaneously contributes to soil fertility because the plant root system leaves behind some of the biologically available nitrogen. As in any symbiosis, both organisms benefit from the interaction: the plant obtains ammonia, and bacteria obtain carbon compounds generated through photosynthesis, as well as a protected niche in which to grow ([\[link\]](#)).



Soybean roots contain (a) nitrogen-fixing nodules. Cells within the nodules are infected with *Bradyrhizobium japonicum*, a rhizobia or “root-loving” bacterium. The bacteria are encased in (b) vesicles inside the cell, as can be seen in this transmission electron micrograph. (credit a: modification of work by USDA; credit b: modification of work by Louisa Howard, Dartmouth Electron Microscope Facility; scale-bar data from Matt Russell)

## Mycorrhizae: The Symbiotic Relationship between Fungi and Roots

A nutrient depletion zone can develop when there is rapid soil solution uptake, low nutrient concentration, low diffusion rate, or low soil moisture. These conditions are very common; therefore, most plants rely on fungi to facilitate the uptake of minerals from the soil. Fungi form symbiotic associations called mycorrhizae with plant roots, in which the fungi actually are integrated into the physical structure of the root. The fungi colonize the living root tissue during active plant growth.

Through mycorrhization, the plant obtains mainly phosphate and other minerals, such as zinc and copper, from the soil. The fungus obtains nutrients, such as sugars, from the plant root ([link](#)). Mycorrhizae help

increase the surface area of the plant root system because hyphae, which are narrow, can spread beyond the nutrient depletion zone. Hyphae can grow into small soil pores that allow access to phosphorus that would otherwise be unavailable to the plant. The beneficial effect on the plant is best observed in poor soils. The benefit to fungi is that they can obtain up to 20 percent of the total carbon accessed by plants. Mycorrhizae functions as a physical barrier to pathogens. It also provides an induction of generalized host defense mechanisms, and sometimes involves production of antibiotic compounds by the fungi.



Root tips proliferate in the presence of mycorrhizal infection, which appears as off-white fuzz in this image. (credit: modification of work by Nilsson et al., BMC Bioinformatics 2005)

There are two types of mycorrhizae: ectomycorrhizae and endomycorrhizae. Ectomycorrhizae form an extensive dense sheath around the roots, called a mantle. Hyphae from the fungi extend from the mantle into the soil, which increases the surface area for water and mineral absorption. This type of mycorrhizae is found in forest trees, especially

conifers, birches, and oaks. Endomycorrhizae, also called arbuscular mycorrhizae, do not form a dense sheath over the root. Instead, the fungal mycelium is embedded within the root tissue. Endomycorrhizae are found in the roots of more than 80 percent of terrestrial plants.

## Nutrients from Other Sources

Some plants cannot produce their own food and must obtain their nutrition from outside sources. This may occur with plants that are parasitic or saprophytic. Some plants are mutualistic symbionts, epiphytes, or insectivorous.

## Plant Parasites

A **parasitic plant** depends on its host for survival. Some parasitic plants have no leaves. An example of this is the dodder ([\[link\]](#)), which has a weak, cylindrical stem that coils around the host and forms suckers. From these suckers, cells invade the host stem and grow to connect with the vascular bundles of the host. The parasitic plant obtains water and nutrients through these connections. The plant is a total parasite (a holoparasite) because it is completely dependent on its host. Other parasitic plants (hemiparasites) are fully photosynthetic and only use the host for water and minerals. There are about 4,100 species of parasitic plants.





The dodder is a holoparasite that penetrates the host's vascular tissue and diverts nutrients for its own growth. Note that the vines of the dodder, which has white flowers, are beige. The dodder has no chlorophyll and cannot produce its own food. (credit: "Lalithamba"/Flickr)

## **Saprophytes**

A **saprophyte** is a plant that does not have chlorophyll and gets its food from dead matter, similar to bacteria and fungi (note that fungi are often called saprophytes, which is incorrect, because fungi are not plants). Plants like these use enzymes to convert organic food materials into simpler forms from which they can absorb nutrients ([\[link\]](#)). Most saprophytes do not directly digest dead matter: instead, they parasitize fungi that digest dead matter, or are mycorrhizal, ultimately obtaining photosynthate from a fungus that derived photosynthate from its host. Saprophytic plants are uncommon; only a few species are described.



Saprophytes, like this Dutchmen's pipe (*Monotropa hypopitys*), obtain their food from dead matter and do not have chlorophyll.  
(credit: modification of work by Iwona Erskine-Kellie)

## Symbionts

A **symbiont** is a plant in a symbiotic relationship, with special adaptations such as mycorrhizae or nodule formation. Fungi also form symbiotic associations with cyanobacteria and green algae (called lichens). Lichens can sometimes be seen as colorful growths on the surface of rocks and trees ([\[link\]](#)). The algal partner (phycobiont) makes food autotrophically, some of which it shares with the fungus; the fungal partner (mycobiont) absorbs water and minerals from the environment, which are made available to the green alga. If one partner was separated from the other, they would both die.



Lichens, which often have symbiotic relationships with other plants, can sometimes be found growing on trees. (credit: "benketaro"/Flickr)

## Epiphytes

An **epiphyte** is a plant that grows on other plants, but is not dependent upon the other plant for nutrition ([link](#)). Epiphytes have two types of roots: clinging aerial roots, which absorb nutrients from humus that accumulates in the crevices of trees; and aerial roots, which absorb moisture from the atmosphere.



These epiphyte plants  
grow in the main  
greenhouse of the *Jardin  
des Plantes* in Paris.

## **Insectivorous Plants**

An **insectivorous** plant has specialized leaves to attract and digest insects. The Venus flytrap is popularly known for its insectivorous mode of nutrition, and has leaves that work as traps ([\[link\]](#)). The minerals it obtains from prey compensate for those lacking in the boggy (low pH) soil of its native North Carolina coastal plains. There are three sensitive hairs in the center of each half of each leaf. The edges of each leaf are covered with long spines. Nectar secreted by the plant attracts flies to the leaf. When a fly touches the sensory hairs, the leaf immediately closes. Next, fluids and enzymes break down the prey and minerals are absorbed by the leaf. Since this plant is popular in the horticultural trade, it is threatened in its original habitat.



A Venus flytrap has specialized leaves to trap insects. (credit: "Selena N. B. H."/Flickr)

## Section Summary

Atmospheric nitrogen is the largest pool of available nitrogen in terrestrial ecosystems. However, plants cannot use this nitrogen because they do not have the necessary enzymes. Biological nitrogen fixation (BNF) is the conversion of atmospheric nitrogen to ammonia. The most important source of BNF is the symbiotic interaction between soil bacteria and legumes. The bacteria form nodules on the legume's roots in which nitrogen fixation takes place. Fungi form symbiotic associations (mycorrhizae) with plants, becoming integrated into the physical structure of the root. Through mycorrhization, the plant obtains minerals from the soil and the fungus obtains photosynthate from the plant root. Ectomycorrhizae form an extensive dense sheath around the root, while endomycorrhizae are embedded within the root tissue. Some plants—parasites, saprophytes, symbionts, epiphytes, and insectivores—have evolved adaptations to obtain their organic or mineral nutrition from various sources.

## Art Connections

**Exercise:****Problem:**

[\[link\]](#) Farmers often rotate corn (a cereal crop) and soy beans (a legume) planting a field with each crop in alternate seasons. What advantage might this crop rotation confer?

---

**Solution:**

[\[link\]](#) Soybeans are able to fix nitrogen in their roots, which are not harvested at the end of the growing season. The belowground nitrogen can be used in the next season by the corn.

**Review Questions****Exercise:****Problem:**

Which process produces an inorganic compound that plants can easily use?

- a. photosynthesis
  - b. nitrogen fixation
  - c. mycorrhization
  - d. Calvin cycle
- 

**Solution:**

B

**Exercise:****Problem:**

Through mycorrhization, a plant obtains important nutrients such as \_\_\_\_\_.

- a. phosphorus, zinc, and copper
- b. phosphorus, zinc, and calcium
- c. nickel, calcium, and zinc
- d. all of the above

---

**Solution:**

A

**Exercise:**

**Problem:**

What term describes a plant that requires nutrition from a living host plant?

- a. parasite
- b. saprophyte
- c. epiphyte
- d. insectivorous

---

**Solution:**

A

**Exercise:**

**Problem:**

What is the term for the symbiotic association between fungi and cyanobacteria?

- a. lichen
- b. mycorrhizae
- c. epiphyte
- d. nitrogen-fixing nodule

---

**Solution:**

A

## Free Response

### Exercise:

#### Problem:

Why is biological nitrogen fixation an environmentally friendly way of fertilizing plants?

---

#### Solution:

Because it is natural and does not require use of a nonrenewable resource, such as natural gas.

### Exercise:

#### Problem:

What is the main difference, from an energy point of view, between photosynthesis and biological nitrogen fixation?

---

#### Solution:

Photosynthesis harvests and stores energy, whereas biological nitrogen fixation requires energy.

### Exercise:

**Problem:** Why is a root nodule a nutritional adaptation of a plant?

---

#### Solution:

A nodule results from the symbiosis between a plant and bacterium. Within nodules, the process of nitrogen fixation allows the plant to obtain nitrogen from the air.



## Glossary

### epiphyte

plant that grows on other plants but is not dependent upon other plants for nutrition

### insectivorous plant

plant that has specialized leaves to attract and digest insects

### nitrogenase

enzyme that is responsible for the reduction of atmospheric nitrogen to ammonia

### nodules

specialized structures that contain *Rhizobia* bacteria where nitrogen fixation takes place

### parasitic plant

plant that is dependent on its host for survival

### rhizobia

soil bacteria that symbiotically interact with legume roots to form nodules and fix nitrogen

### saprophyte

plant that does not have chlorophyll and gets its food from dead matter

### symbiont

plant in a symbiotic relationship with bacteria or fungi

## Introduction

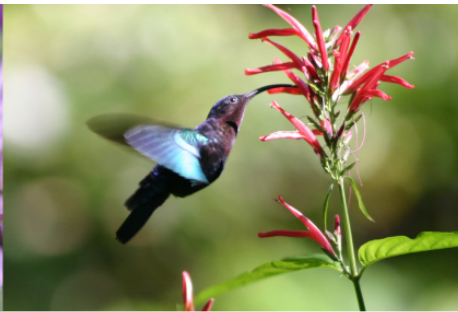
class="introduction"

Plants that reproduce sexually often achieve fertilization with the help of pollinators such as (a) bees, (b) birds, and (c) butterflies.

(credit a: modification of work by John Severns; credit b: modification of work by Charles J. Sharp; credit c: modification of work by "Galawebdesign"/Flickr )



(a)



(b)



(c)

Plants have evolved different reproductive strategies for the continuation of their species. Some plants reproduce sexually, and others asexually, in contrast to animal species, which rely almost exclusively on sexual reproduction. Plant sexual reproduction usually depends on pollinating agents, while asexual reproduction is independent of these agents. Flowers are often the showiest or most strongly scented part of plants. With their bright colors, fragrances, and interesting shapes and sizes, flowers attract

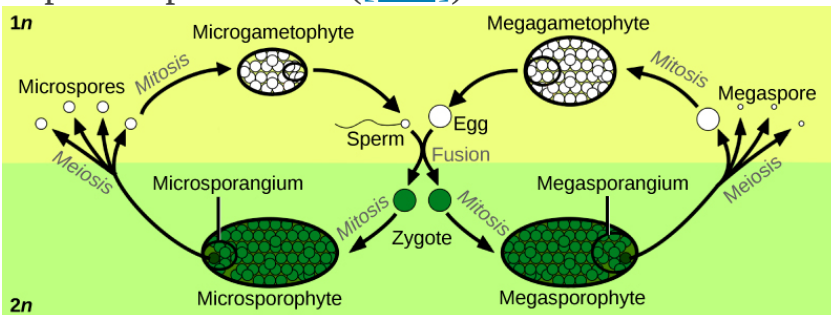
insects, birds, and animals to serve their pollination needs. Other plants pollinate via wind or water; still others self-pollinate.

## Reproductive Development and Structure

By the end of this section, you will be able to:

- Describe the two stages of a plant's lifecycle
- Compare and contrast male and female gametophytes and explain how they form in angiosperms
- Describe the reproductive structures of a plant
- Describe the components of a complete flower
- Describe the development of microsporangium and megasporangium in gymnosperms

Sexual reproduction takes place with slight variations in different groups of plants. Plants have two distinct stages in their lifecycle: the gametophyte stage and the sporophyte stage. The haploid **gametophyte** produces the male and female gametes by mitosis in distinct multicellular structures. Fusion of the male and female gametes forms the diploid zygote, which develops into the **sporophyte**. After reaching maturity, the diploid sporophyte produces spores by meiosis, which in turn divide by mitosis to produce the haploid gametophyte. The new gametophyte produces gametes, and the cycle continues. This is the alternation of generations, and is typical of plant reproduction ([\[link\]](#)).



The alternation of generations in angiosperms is depicted in this diagram. (credit: modification of work by Peter Coxhead)

The life cycle of higher plants is dominated by the sporophyte stage, with the gametophyte borne on the sporophyte. In ferns, the gametophyte is free-

living and very distinct in structure from the diploid sporophyte. In bryophytes, such as mosses, the haploid gametophyte is more developed than the sporophyte.

During the vegetative phase of growth, plants increase in size and produce a shoot system and a root system. As they enter the reproductive phase, some of the branches start to bear flowers. Many flowers are borne singly, whereas some are borne in clusters. The flower is borne on a stalk known as a receptacle. Flower shape, color, and size are unique to each species, and are often used by taxonomists to classify plants.

## Sexual Reproduction in Angiosperms

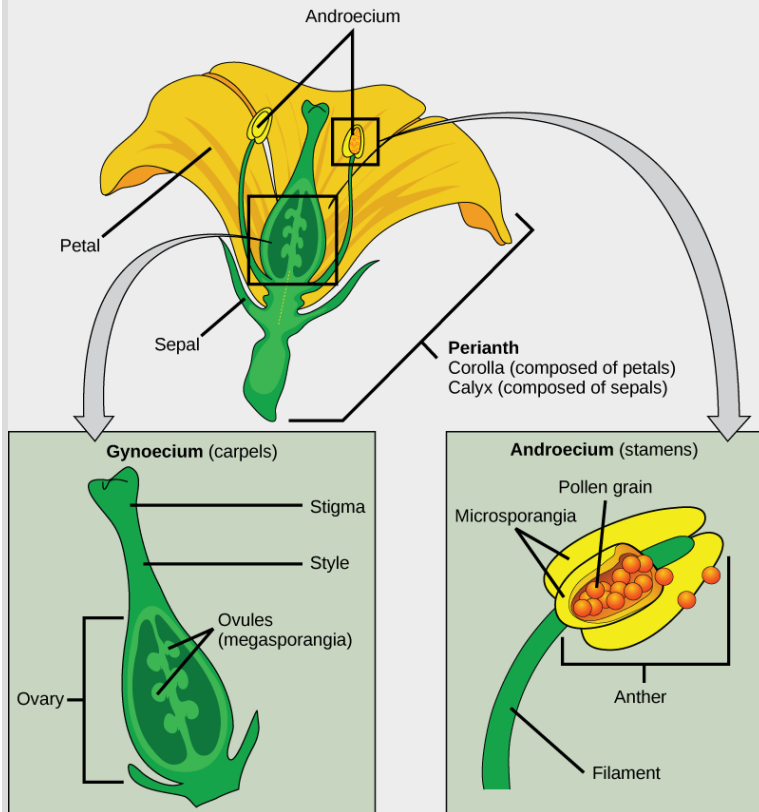
The lifecycle of angiosperms follows the alternation of generations explained previously. The haploid gametophyte alternates with the diploid sporophyte during the sexual reproduction process of angiosperms. Flowers contain the plant's reproductive structures.

### Flower Structure

A typical flower has four main parts—or whorls—known as the calyx, corolla, androecium, and gynoecium ([link](#)). The outermost whorl of the flower has green, leafy structures known as sepals. The sepals, collectively called the calyx, help to protect the unopened bud. The second whorl is comprised of petals—usually, brightly colored—collectively called the corolla. The number of sepals and petals varies depending on whether the plant is a monocot or dicot. In monocots, petals usually number three or multiples of three; in dicots, the number of petals is four or five, or multiples of four and five. Together, the calyx and corolla are known as the **perianth**. The third whorl contains the male reproductive structures and is known as the androecium. The **androecium** has stamens with anthers that contain the microsporangia. The innermost group of structures in the flower is the **gynoecium**, or the female reproductive component(s). The carpel is the individual unit of the gynoecium and has a stigma, style, and ovary. A flower may have one or multiple carpels.

**Note:**

**Art Connection**



The four main parts of the flower are the calyx, corolla, androecium, and gynoecium. The androecium is the sum of all the male reproductive organs, and the gynoecium is the sum of the female reproductive organs. (credit: modification of work by Mariana Ruiz Villareal)

If the anther is missing, what type of reproductive structure will the flower be unable to produce? What term is used to describe an incomplete flower lacking the androecium? What term describes an incomplete flower lacking a gynoecium?

If all four whorls (the calyx, corolla, androecium, and gynoecium) are present, the flower is described as complete. If any of the four parts is missing, the flower is known as incomplete. Flowers that contain both an androecium and a gynoecium are called perfect, androgynous or hermaphrodites. There are two types of incomplete flowers: staminate flowers contain only an androecium, and carpellate flowers have only a gynoecium ([\[link\]](#)).

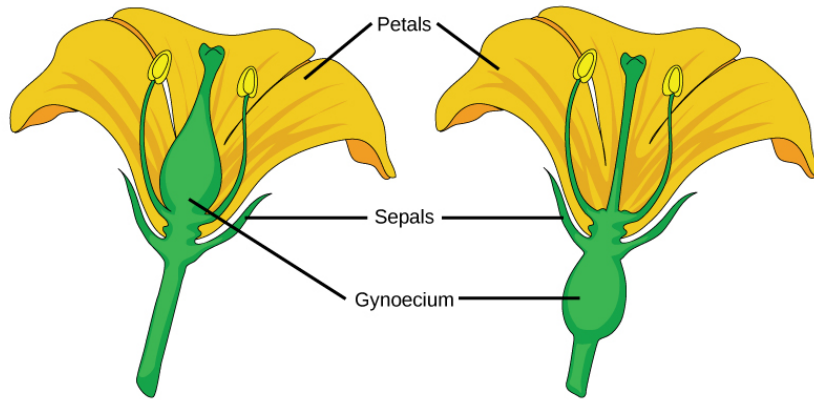


The corn plant has both staminate (male) and carpellate (female) flowers. Staminate flowers, which are clustered in the tassel at the tip of the stem, produce pollen grains. Carpellate flower are clustered in

the immature ears. Each strand of silk is a stigma. The corn kernels are seeds that develop on the ear after fertilization. Also shown is the lower stem and root.

If both male and female flowers are borne on the same plant, the species is called monoecious (meaning “one home”): examples are corn and pea. Species with male and female flowers borne on separate plants are termed dioecious, or “two homes,” examples of which are *C. papaya* and *Cannabis*. The ovary, which may contain one or multiple ovules, may be placed above other flower parts, which is referred to as superior; or, it may be placed below the other flower parts, referred to as inferior ([\[link\]](#)).





(a) Superior flower



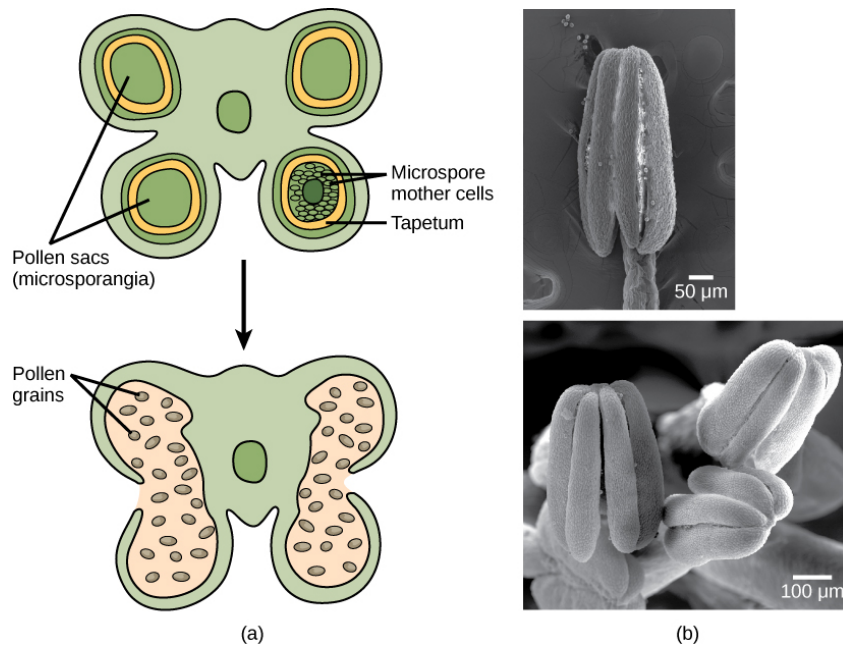
(b) Inferior flower

The (a) lily is a superior flower, which has the ovary above the other flower parts. (b) Fuchsia is an inferior flower, which has the ovary beneath other flower parts. (credit a photo: modification of work by Benjamin Zwittnig; credit b photo: modification of work by "Koshy Koshy"/Flickr)

## Male Gametophyte (The Pollen Grain)

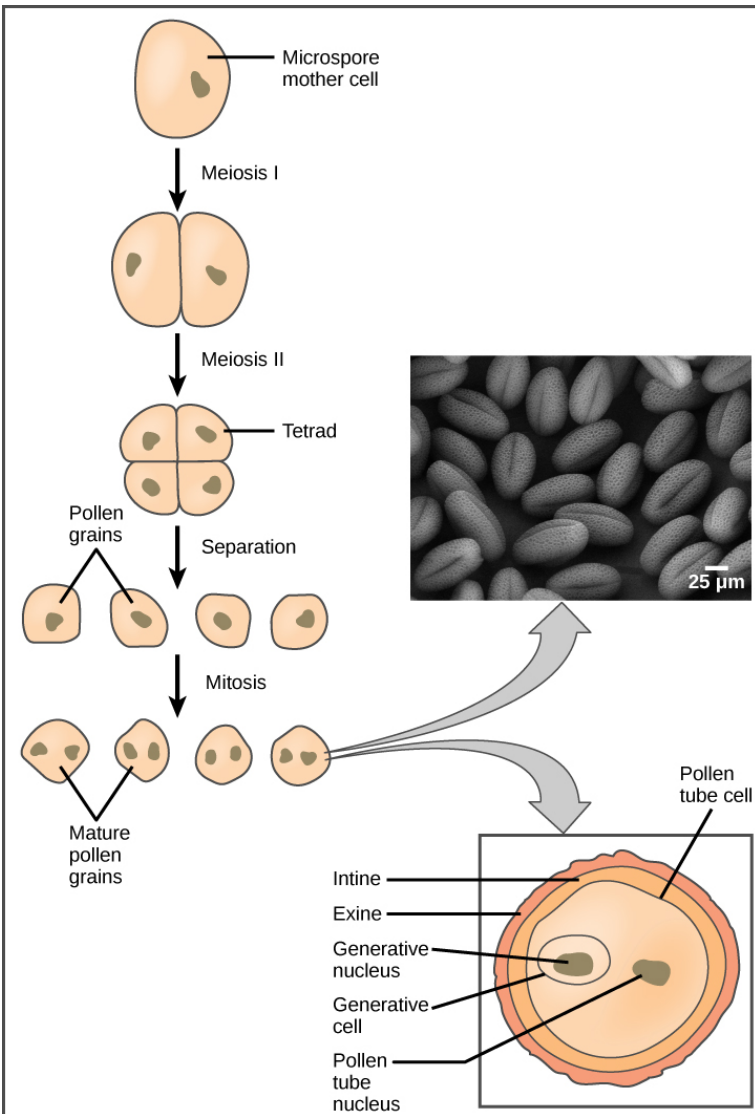
The male gametophyte develops and reaches maturity in an immature anther. In a plant's male reproductive organs, development of pollen takes

place in a structure known as the **microsporangium** ([link](#)). The microsporangia, which are usually bi-lobed, are pollen sacs in which the microspores develop into pollen grains. These are found in the anther, which is at the end of the stamen—the long filament that supports the anther.



Shown is (a) a cross section of an anther at two developmental stages. The immature anther (top) contains four microsporangia, or pollen sacs. Each microsporangium contains hundreds of microspore mother cells that will each give rise to four pollen grains. The tapetum supports the development and maturation of the pollen grains. Upon maturation of the pollen (bottom), the pollen sac walls split open and the pollen grains (male gametophytes) are released. (b) In these scanning electron micrographs, pollen sacs are ready to burst, releasing their grains. (credit b: modification of work by Robert R. Wise; scale-bar data from Matt Russell)

Within the microsporangium, the microspore mother cell divides by meiosis to give rise to four microspores, each of which will ultimately form a pollen grain ([link](#)). An inner layer of cells, known as the tapetum, provides nutrition to the developing microspores and contributes key components to the pollen wall. Mature pollen grains contain two cells: a generative cell and a pollen tube cell. The generative cell is contained within the larger pollen tube cell. Upon germination, the tube cell forms the pollen tube through which the generative cell migrates to enter the ovary. During its transit inside the pollen tube, the generative cell divides to form two male gametes (sperm cells). Upon maturity, the microsporangia burst, releasing the pollen grains from the anther.



Pollen develops from the microspore mother cells. The mature pollen grain is composed of two cells: the pollen tube cell and the generative cell, which is inside the tube cell. The pollen grain has two coverings: an inner layer (intine) and an outer layer (exine). The inset scanning electron micrograph shows *Arabidopsis lyrata* pollen grains. (credit “pollen micrograph”: modification of work by Robert R. Wise; scale-bar data from Matt Russell)

Each pollen grain has two coverings: the **exine** (thicker, outer layer) and the **intine** ([\[link\]](#)). The exine contains sporopollenin, a complex waterproofing substance supplied by the tapetal cells. Sporopollenin allows the pollen to survive under unfavorable conditions and to be carried by wind, water, or biological agents without undergoing damage.

### Female Gametophyte (The Embryo Sac)

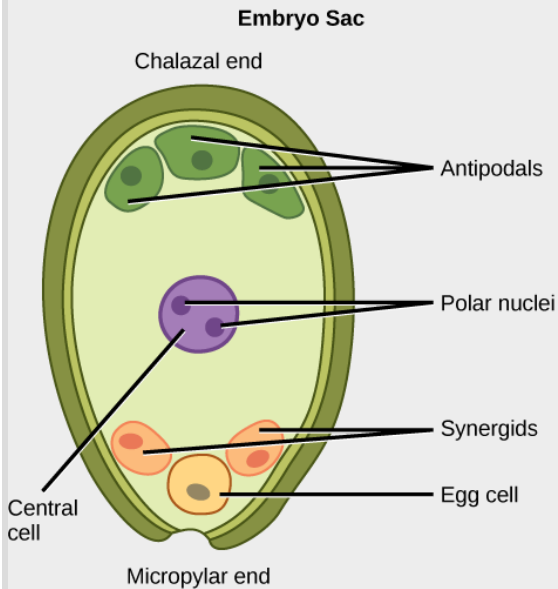
While the details may vary between species, the overall development of the female gametophyte has two distinct phases. First, in the process of **megasporogenesis**, a single cell in the diploid **megasporangium**—an area of tissue in the ovules—undergoes meiosis to produce four megaspores, only one of which survives. During the second phase, **megagametogenesis**, the surviving haploid megaspore undergoes mitosis to produce an eight-nucleate, seven-cell female gametophyte, also known as the megagametophyte or embryo sac. Two of the nuclei—the **polar nuclei**—move to the equator and fuse, forming a single, diploid central cell. This central cell later fuses with a sperm to form the triploid endosperm. Three nuclei position themselves on the end of the embryo sac opposite the micropyle and develop into the **antipodal** cells, which later degenerate. The nucleus closest to the micropyle becomes the female gamete, or egg cell, and the two adjacent nuclei develop into **synergid** cells ([\[link\]](#)). The synergids help guide the pollen tube for successful fertilization, after which they disintegrate. Once fertilization is complete, the resulting diploid zygote develops into the embryo, and the fertilized ovule forms the other tissues of the seed.

A double-layered integument protects the megasporangium and, later, the embryo sac. The integument will develop into the seed coat after fertilization and protect the entire seed. The ovule wall will become part of the fruit. The integuments, while protecting the megasporangium, do not enclose it completely, but leave an opening called the **micropyle**. The

micropyle allows the pollen tube to enter the female gametophyte for fertilization.

**Note:**

**Art Connection**



As shown in this diagram of the embryo sac in angiosperms, the ovule is covered by integuments and has an opening called a micropyle. Inside the embryo sac are three antipodal cells, two synergids, a central cell, and the egg cell.

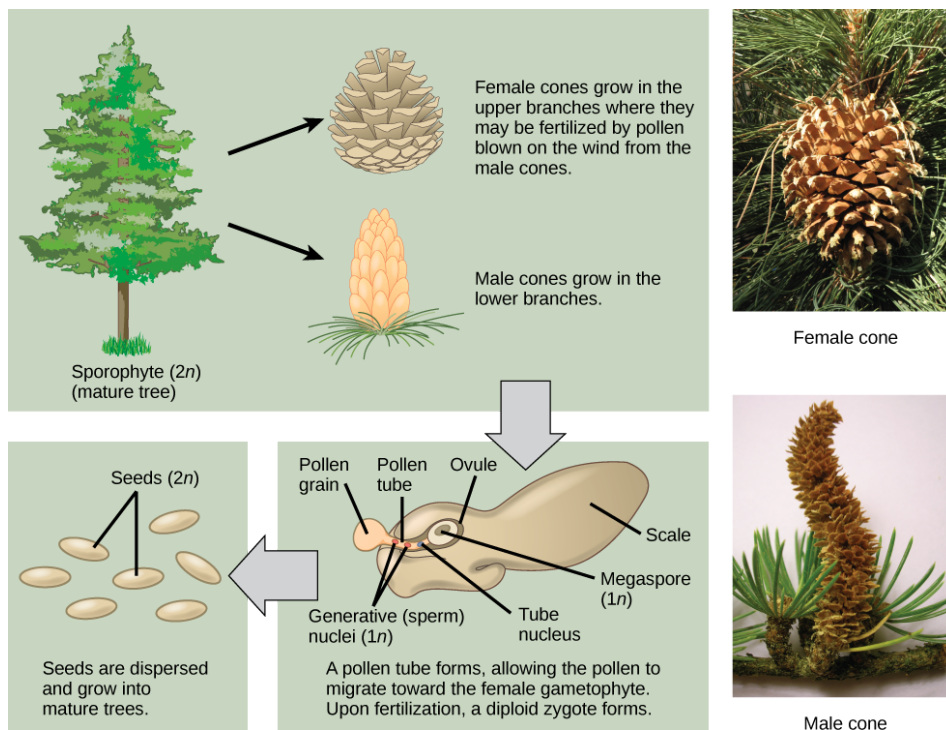
An embryo sac is missing the synergids. What specific impact would you expect this to have on fertilization?

- a. The pollen tube will be unable to form.
- b. The pollen tube will form but will not be guided toward the egg.

- c. Fertilization will not occur because the synergid is the egg.
- d. Fertilization will occur but the embryo will not be able to grow.

## Sexual Reproduction in Gymnosperms

As with angiosperms, the lifecycle of a gymnosperm is also characterized by alternation of generations. In conifers such as pines, the green leafy part of the plant is the sporophyte, and the cones contain the male and female gametophytes ([\[link\]](#)). The female cones are larger than the male cones and are positioned towards the top of the tree; the small, male cones are located in the lower region of the tree. Because the pollen is shed and blown by the wind, this arrangement makes it difficult for a gymnosperm to self-pollinate.



This image shows the life cycle of a conifer. Pollen from male cones blows up into upper branches, where

it fertilizes female cones. Examples are shown of female and male cones. (credit “female”: modification of work by “Geographer”/Wikimedia Commons; credit “male”: modification of work by Roger Griffith)

## Male Gametophyte

A male cone has a central axis on which bracts, a type of modified leaf, are attached. The bracts are known as **microsporophylls** ([\[link\]](#)) and are the sites where microspores will develop. The microspores develop inside the microsporangium. Within the microsporangium, cells known as microsporocytes divide by meiosis to produce four haploid microspores. Further mitosis of the microspore produces two nuclei: the generative nucleus, and the tube nucleus. Upon maturity, the male gametophyte (pollen) is released from the male cones and is carried by the wind to land on the female cone.

### Note:

Link to Learning



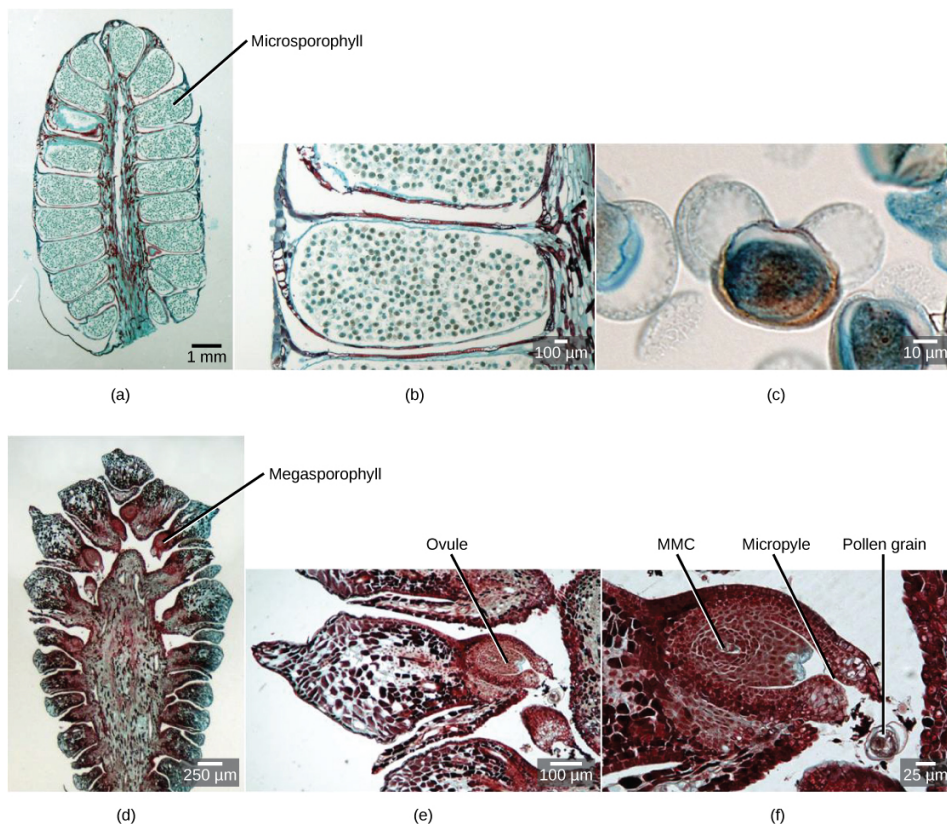
Watch this video to see a cedar releasing its pollen in the wind.

[https://www.openstaxcollege.org/l/pollen\\_release](https://www.openstaxcollege.org/l/pollen_release)



## Female Gametophyte

The female cone also has a central axis on which bracts known as **megasporophylls** ([\[link\]](#)) are present. In the female cone, megaspore mother cells are present in the megasporangium. The megaspore mother cell divides by meiosis to produce four haploid megaspores. One of the megaspores divides to form the multicellular female gametophyte, while the others divide to form the rest of the structure. The female gametophyte is contained within a structure called the archegonium.



These series of micrographs shows male and female gymnosperm gametophytes. (a) This male cone, shown in cross section, has approximately 20 microsporophylls, each of which produces hundreds of male gametophytes (pollen grains). (b) Pollen grains are visible in this single microsporophyll. (c) This micrograph shows an individual pollen grain. (d)

This cross section of a female cone shows portions of about 15 megasporophylls. (e) The ovule can be seen in this single megasporophyll. (f) Within this single ovule are the megaspore mother cell (MMC), micropyle, and a pollen grain. (credit: modification of work by Robert R. Wise; scale-bar data from Matt Russell)

## **Reproductive Process**

Upon landing on the female cone, the tube cell of the pollen forms the pollen tube, through which the generative cell migrates towards the female gametophyte through the micropyle. It takes approximately one year for the pollen tube to grow and migrate towards the female gametophyte. The male gametophyte containing the generative cell splits into two sperm nuclei, one of which fuses with the egg, while the other degenerates. After fertilization of the egg, the diploid zygote is formed, which divides by mitosis to form the embryo. The scales of the cones are closed during development of the seed. The seed is covered by a seed coat, which is derived from the female sporophyte. Seed development takes another one to two years. Once the seed is ready to be dispersed, the bracts of the female cones open to allow the dispersal of seed; no fruit formation takes place because gymnosperm seeds have no covering.

## **Angiosperms versus Gymnosperms**

Gymnosperm reproduction differs from that of angiosperms in several ways ([link](#)). In angiosperms, the female gametophyte exists in an enclosed structure—the ovule—which is within the ovary; in gymnosperms, the female gametophyte is present on exposed bracts of the female cone. Double fertilization is a key event in the lifecycle of angiosperms, but is completely absent in gymnosperms. The male and female gametophyte structures are present on separate male and female cones in gymnosperms,

whereas in angiosperms, they are a part of the flower. Lastly, wind plays an important role in pollination in gymnosperms because pollen is blown by the wind to land on the female cones. Although many angiosperms are also wind-pollinated, animal pollination is more common.



(a)



(b)

(a) Angiosperms are flowering plants, and include grasses, herbs, shrubs and most deciduous trees, while (b) gymnosperms are conifers. Both produce seeds but have different reproductive strategies. (credit a: modification of work by Wendy Cutler; credit b: modification of work by Lews Castle UHI)

**Note:**

[Link to Learning](#)



View an animation of the double fertilization process of angiosperms.  
<https://www.openstaxcollege.org/l/angiosperms>

## Section Summary

The flower contains the reproductive structures of a plant. All complete flowers contain four whorls: the calyx, corolla, androecium, and gynoecium. The stamens are made up of anthers, in which pollen grains are produced, and a supportive strand called the filament. The pollen contains two cells—a generative cell and a tube cell—and is covered by two layers called the intine and the exine. The carpels, which are the female reproductive structures, consist of the stigma, style, and ovary. The female gametophyte is formed from mitotic divisions of the megaspore, forming an eight-nuclei ovule sac. This is covered by a layer known as the integument. The integument contains an opening called the micropyle, through which the pollen tube enters the embryo sac.

The diploid sporophyte of angiosperms and gymnosperms is the conspicuous and long-lived stage of the life cycle. The sporophytes differentiate specialized reproductive structures called sporangia, which are dedicated to the production of spores. The microsporangium contains microspore mother cells, which divide by meiosis to produce haploid microspores. The microspores develop into male gametophytes that are released as pollen. The megasporangium contains megaspore mother cells, which divide by meiosis to produce haploid megaspores. A megaspore develops into a female gametophyte containing a haploid egg. A new diploid sporophyte is formed when a male gamete from a pollen grain enters the ovule sac and fertilizes this egg.

## Art Connections

### Exercise:

#### Problem:

[\[link\]](#) If the anther is missing, what type of reproductive structure will the flower be unable to produce? What term is used to describe a flower that is normally lacking the androecium? What term describes a flower lacking a gynoecium?

---

#### Solution:

[\[link\]](#) Pollen (or sperm); carpellate; staminate.

### Exercise:

#### Problem:

[\[link\]](#) An embryo sac is missing the synergids. What specific impact would you expect this to have on fertilization?

- a. The pollen tube will be unable to form.
  - b. The pollen tube will form but will not be guided toward the egg.
  - c. Fertilization will not occur because the synergid is the egg.
  - d. Fertilization will occur but the embryo will not be able to grow.
- 

#### Solution:

[\[link\]](#) B: The pollen tube will form but will not be guided toward the egg.

## Review Questions

### Exercise:

**Problem:**

In a plant's male reproductive organs, development of pollen takes place in a structure known as the \_\_\_\_\_.

- a. stamen
- b. microsporangium
- c. anther
- d. tapetum

---

**Solution:**

B

**Exercise:****Problem:**

The stamen consists of a long stalk called the filament that supports the \_\_\_\_\_.

- a. stigma
- b. sepal
- c. style
- d. anther

---

**Solution:**

D

**Exercise:**

**Problem:** The \_\_\_\_\_ are collectively called the calyx.

- a. sepals
- b. petals
- c. tepals

d. stamens

---

**Solution:**

A

**Exercise:**

**Problem:** The pollen lands on which part of the flower?

- a. stigma
- b. style
- c. ovule
- d. integument

---

**Solution:**

A

**Free Response**

**Exercise:**

**Problem:** Describe the reproductive organs inside a flower.

---

**Solution:**

Inside the flower are the reproductive organs of the plant. The stamen is the male reproductive organ. Pollen is produced in the stamen. The carpel is the female reproductive organ. The ovary is the swollen base of the carpel where ovules are found. Not all flowers have every one of the four parts.

**Exercise:**

**Problem:**

Describe the two-stage lifecycle of plants: the gametophyte stage and the sporophyte stage.

---

**Solution:**

Plants have two distinct phases in their lifecycle: the gametophyte stage and the sporophyte stage. In the gametophyte stage, when reproductive cells undergo meiosis and produce haploid cells called spores, the gametophyte stage begins. Spores divide by cell division to form plant structures of an entirely new plant. The cells in these structures or plants are haploid. Some of these cells undergo cell division and form sex cells. Fertilization, the joining of haploid sex cells, begins the sporophyte stage. Cells formed in this stage have the diploid number of chromosomes. Meiosis in some of these cells forms spores, and the cycle begins again: a process known as alternation of generations.

**Exercise:**

**Problem:** Describe the four main parts, or whorls, of a flower.

---

**Solution:**

A typical flower has four main parts, or whorls: the calyx, corolla, androecium, and gynoecium. The outermost whorl of the flower has green, leafy structures known as sepals, which are collectively called the calyx. It helps to protect the unopened bud. The second whorl is made up of brightly colored petals that are known collectively as the corolla. The third whorl is the male reproductive structure known as the androecium. The androecium has stamens, which have anthers on a stalk or filament. Pollen grains are borne on the anthers. The gynoecium is the female reproductive structure. The carpel is the individual structure of the gynoecium and has a stigma, the stalk or style, and the ovary.

**Exercise:**



**Problem:**

Discuss the differences between a complete flower and an incomplete flower.

---

**Solution:**

If all four whorls of a flower are present, it is a complete flower. If any of the four parts is missing, it is known as incomplete. Flowers that contain both an androecium and gynoecium are called androgynous or hermaphrodites. Those that contain only an androecium are known as staminate flowers, and those that have only carpels are known as carpellate. If both male and female flowers are borne on the same plant, it is called monoecious, while plants with male and female flowers on separate plants are termed dioecious.

**Glossary**

androecium

sum of all the stamens in a flower

antipodals

the three cells away from the micropyle

exine

outermost covering of pollen

gametophyte

multicellular stage of the plant that gives rise to haploid gametes or spores

gynoecium

the sum of all the carpels in a flower

intine

inner lining of the pollen

megagametogenesis

second phase of female gametophyte development, during which the surviving haploid megaspore undergoes mitosis to produce an eight-nucleate, seven-cell female gametophyte, also known as the megagametophyte or embryo sac.

megasporangium

tissue found in the ovary that gives rise to the female gamete or egg

megasporogenesis

first phase of female gametophyte development, during which a single cell in the diploid megasporangium undergoes meiosis to produce four megaspores, only one of which survives

megasporophyll

bract (a type of modified leaf) on the central axis of a female gametophyte

micropyle

opening on the ovule sac through which the pollen tube can gain entry

microsporangium

tissue that gives rise to the microspores or the pollen grain

microsporophyll

central axis of a male cone on which bracts (a type of modified leaf) are attached

perianth

(also, petal or sepal) part of the flower consisting of the calyx and/or corolla; forms the outer envelope of the flower

polar nuclei

found in the ovule sac; fusion with one sperm cell forms the endosperm

sporophyte

multicellular diploid stage in plants that is formed after the fusion of male and female gametes

synergid

type of cell found in the ovule sac that secretes chemicals to guide the pollen tube towards the egg

## Pollination and Fertilization

By the end of this section, you will be able to:

- Describe what must occur for plant fertilization
- Explain cross-pollination and the ways in which it takes place
- Describe the process that leads to the development of a seed
- Define double fertilization

In angiosperms, **pollination** is defined as the placement or transfer of pollen from the anther to the stigma of the same flower or another flower. In gymnosperms, pollination involves pollen transfer from the male cone to the female cone. Upon transfer, the pollen germinates to form the pollen tube and the sperm for fertilizing the egg. Pollination has been well studied since the time of Gregor Mendel. Mendel successfully carried out self- as well as cross-pollination in garden peas while studying how characteristics were passed on from one generation to the next. Today's crops are a result of plant breeding, which employs artificial selection to produce the present-day cultivars. A case in point is today's corn, which is a result of years of breeding that started with its ancestor, teosinte. The teosinte that the ancient Mayans originally began cultivating had tiny seeds—vastly different from today's relatively giant ears of corn. Interestingly, though these two plants appear to be entirely different, the genetic difference between them is miniscule.

Pollination takes two forms: self-pollination and cross-pollination. **Self-pollination** occurs when the pollen from the anther is deposited on the stigma of the same flower, or another flower on the same plant. **Cross-pollination** is the transfer of pollen from the anther of one flower to the stigma of another flower on a different individual of the same species. Self-pollination occurs in flowers where the stamen and carpel mature at the same time, and are positioned so that the pollen can land on the flower's stigma. This method of pollination does not require an investment from the plant to provide nectar and pollen as food for pollinators.

### Note:

[Link to Learning](#)



Explore this [interactive website](#) to review self-pollination and cross-pollination.

Living species are designed to ensure survival of their progeny; those that fail become extinct. Genetic diversity is therefore required so that in changing environmental or stress conditions, some of the progeny can survive. Self-pollination leads to the production of plants with less genetic diversity, since genetic material from the same plant is used to form gametes, and eventually, the zygote. In contrast, cross-pollination—or out-crossing—leads to greater genetic diversity because the microgametophyte and megagametophyte are derived from different plants.

Because cross-pollination allows for more genetic diversity, plants have developed many ways to avoid self-pollination. In some species, the pollen and the ovary mature at different times. These flowers make self-pollination nearly impossible. By the time pollen matures and has been shed, the stigma of this flower is mature and can only be pollinated by pollen from another flower. Some flowers have developed physical features that prevent self-pollination. The primrose is one such flower. Primroses have evolved two flower types with differences in anther and stigma length: the pin-eyed flower has anthers positioned at the pollen tube's halfway point, and the thrum-eyed flower's stigma is likewise located at the halfway point. Insects easily cross-pollinate while seeking the nectar at the bottom of the pollen tube. This phenomenon is also known as heterostyly. Many plants, such as cucumber, have male and female flowers located on different parts of the plant, thus making self-pollination difficult. In yet other species, the male and female flowers are borne on different plants (dioecious). All of these are barriers to self-pollination; therefore, the plants depend on pollinators to transfer pollen. The majority of pollinators are biotic agents such as insects

(like bees, flies, and butterflies), bats, birds, and other animals. Other plant species are pollinated by abiotic agents, such as wind and water.

**Note:**

**Everyday Connection**

**Incompatibility Genes in Flowers**

In recent decades, incompatibility genes—which prevent pollen from germinating or growing into the stigma of a flower—have been discovered in many angiosperm species. If plants do not have compatible genes, the pollen tube stops growing. Self-incompatibility is controlled by the S (sterility) locus. Pollen tubes have to grow through the tissue of the stigma and style before they can enter the ovule. The carpel is selective in the type of pollen it allows to grow inside. The interaction is primarily between the pollen and the stigma epidermal cells. In some plants, like cabbage, the pollen is rejected at the surface of the stigma, and the unwanted pollen does not germinate. In other plants, pollen tube germination is arrested after growing one-third the length of the style, leading to pollen tube death. Pollen tube death is due either to apoptosis (programmed cell death) or to degradation of pollen tube RNA. The degradation results from the activity of a ribonuclease encoded by the S locus. The ribonuclease is secreted from the cells of the style in the extracellular matrix, which lies alongside the growing pollen tube. In summary, self-incompatibility is a mechanism that prevents self-fertilization in many flowering plant species. The working of this self-incompatibility mechanism has important consequences for plant breeders because it inhibits the production of inbred and hybrid plants.

## **Pollination by Insects**

Bees are perhaps the most important pollinator of many garden plants and most commercial fruit trees ([\[link\]](#)). The most common species of bees are bumblebees and honeybees. Since bees cannot see the color red, bee-pollinated flowers usually have shades of blue, yellow, or other colors. Bees collect energy-rich pollen or nectar for their survival and energy needs.

They visit flowers that are open during the day, are brightly colored, have a strong aroma or scent, and have a tubular shape, typically with the presence of a nectar guide. A **nectar guide** includes regions on the flower petals that are visible only to bees, and not to humans; it helps to guide bees to the center of the flower, thus making the pollination process more efficient. The pollen sticks to the bees' fuzzy hair, and when the bee visits another flower, some of the pollen is transferred to the second flower. Recently, there have been many reports about the declining population of honeybees. Many flowers will remain unpollinated and not bear seed if honeybees disappear. The impact on commercial fruit growers could be devastating.



Insects, such as bees, are  
important agents of pollination.  
(credit: modification of work by  
Jon Sullivan)

Many flies are attracted to flowers that have a decaying smell or an odor of rotting flesh. These flowers, which produce nectar, usually have dull colors, such as brown or purple. They are found on the corpse flower or voodoo lily (*Amorphophallus*), dragon arum (*Dracunculus*), and carrion flower (*Stapleia*, *Rafflesia*). The nectar provides energy, whereas the pollen

provides protein. Wasps are also important insect pollinators, and pollinate many species of figs.

Butterflies, such as the monarch, pollinate many garden flowers and wildflowers, which usually occur in clusters. These flowers are brightly colored, have a strong fragrance, are open during the day, and have nectar guides to make access to nectar easier. The pollen is picked up and carried on the butterfly's limbs. Moths, on the other hand, pollinate flowers during the late afternoon and night. The flowers pollinated by moths are pale or white and are flat, enabling the moths to land. One well-studied example of a moth-pollinated plant is the yucca plant, which is pollinated by the yucca moth. The shape of the flower and moth have adapted in such a way as to allow successful pollination. The moth deposits pollen on the sticky stigma for fertilization to occur later. The female moth also deposits eggs into the ovary. As the eggs develop into larvae, they obtain food from the flower and developing seeds. Thus, both the insect and flower benefit from each other in this symbiotic relationship. The corn earworm moth and Gaura plant have a similar relationship ([link](#)).



A corn earworm sips nectar from a night-blooming Gaura plant.  
(credit: Juan Lopez, USDA ARS)



## **Pollination by Bats**

In the tropics and deserts, bats are often the pollinators of nocturnal flowers such as agave, guava, and morning glory. The flowers are usually large and white or pale-colored; thus, they can be distinguished from the dark surroundings at night. The flowers have a strong, fruity, or musky fragrance and produce large amounts of nectar. They are naturally large and wide-mouthed to accommodate the head of the bat. As the bats seek the nectar, their faces and heads become covered with pollen, which is then transferred to the next flower.

## **Pollination by Birds**

Many species of small birds, such as the hummingbird ([link](#)) and sun birds, are pollinators for plants such as orchids and other wildflowers. Flowers visited by birds are usually sturdy and are oriented in such a way as to allow the birds to stay near the flower without getting their wings entangled in the nearby flowers. The flower typically has a curved, tubular shape, which allows access for the bird's beak. Brightly colored, odorless flowers that are open during the day are pollinated by birds. As a bird seeks energy-rich nectar, pollen is deposited on the bird's head and neck and is then transferred to the next flower it visits. Botanists have been known to determine the range of extinct plants by collecting and identifying pollen from 200-year-old bird specimens from the same site.



Hummingbirds have adaptations that allow them to reach the nectar of certain tubular flowers. (credit: Lori Branham)

## **Pollination by Wind**

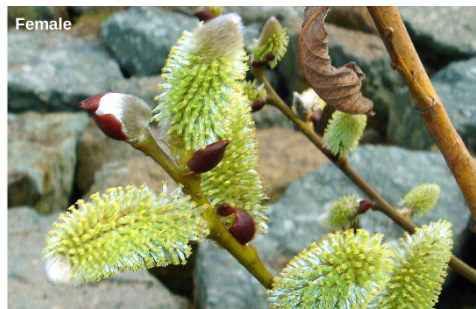
Most species of conifers, and many angiosperms, such as grasses, maples and oaks, are pollinated by wind. Pine cones are brown and unscented, while the flowers of wind-pollinated angiosperm species are usually green, small, may have small or no petals, and produce large amounts of pollen. Unlike the typical insect-pollinated flowers, flowers adapted to pollination by wind do not produce nectar or scent. In wind-pollinated species, the microsporangia hang out of the flower, and, as the wind blows, the lightweight pollen is carried with it ([link](#)). The flowers usually emerge early in the spring, before the leaves, so that the leaves do not block the movement of the wind. The pollen is deposited on the exposed feathery stigma of the flower ([link](#)).



A person knocks pollen from a pine tree.



(a)



(b)

These male (a) and female (b) catkins are from the goat willow tree (*Salix caprea*). Note how both structures are light and feathery to better disperse and catch the wind-blown pollen.

## Pollination by Water

Some weeds, such as Australian sea grass and pond weeds, are pollinated by water. The pollen floats on water, and when it comes into contact with the flower, it is deposited inside the flower.

### Note:

Evolution Connection

### Pollination by Deception

Orchids are highly valued flowers, with many rare varieties ([\[link\]](#)). They grow in a range of specific habitats, mainly in the tropics of Asia, South America, and Central America. At least 25,000 species of orchids have been identified.



Certain orchids use food deception or sexual deception to attract pollinators. Shown here is a bee orchid (*Ophrys apifera*). (credit: David Evans)

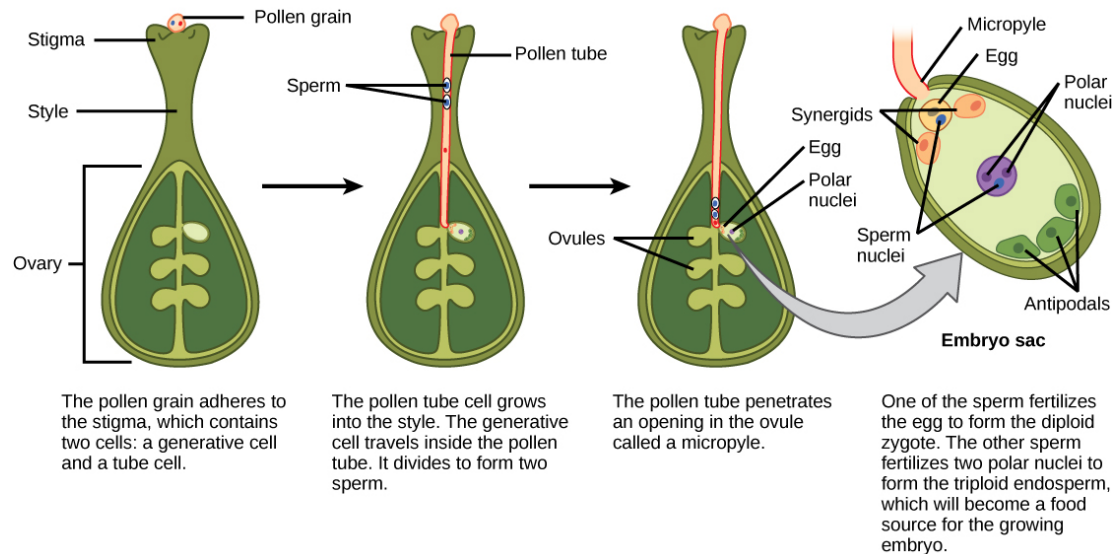
Flowers often attract pollinators with food rewards, in the form of nectar. However, some species of orchid are an exception to this standard: they have evolved different ways to attract the desired pollinators. They use a method known as food deception, in which bright colors and perfumes are offered, but no food. *Anacamptis morio*, commonly known as the green-winged orchid, bears bright purple flowers and emits a strong scent. The bumblebee, its main pollinator, is attracted to the flower because of the strong scent—which usually indicates food for a bee—and in the process, picks up the pollen to be transported to another flower. Other orchids use sexual deception. *Chiloglottis trapeziformis* emits a compound that smells the same as the pheromone emitted by a female wasp to attract male wasps. The male wasp is attracted to the scent, lands on the orchid flower, and in the process, transfers pollen. Some orchids, like the Australian hammer orchid, use scent as well as visual trickery in yet another sexual deception strategy to attract wasps. The flower of this orchid mimics the appearance of a female wasp and emits a pheromone. The male wasp tries to mate with what appears to be a female wasp, and in the process, picks up pollen, which it then transfers to the next counterfeit mate.

## Double Fertilization

After pollen is deposited on the stigma, it must germinate and grow through the style to reach the ovule. The microspores, or the pollen, contain two cells: the pollen tube cell and the generative cell. The pollen tube cell grows into a pollen tube through which the generative cell travels. The germination of the pollen tube requires water, oxygen, and certain chemical signals. As it travels through the style to reach the embryo sac, the pollen tube's growth is supported by the tissues of the style. In the meantime, if the generative cell has not already split into two cells, it now divides to form two sperm cells. The pollen tube is guided by the chemicals secreted by the synergids present in the embryo sac, and it enters the ovule sac through the micropyle. Of the two sperm cells, one sperm fertilizes the egg cell, forming a diploid zygote; the other sperm fuses with the two polar nuclei,



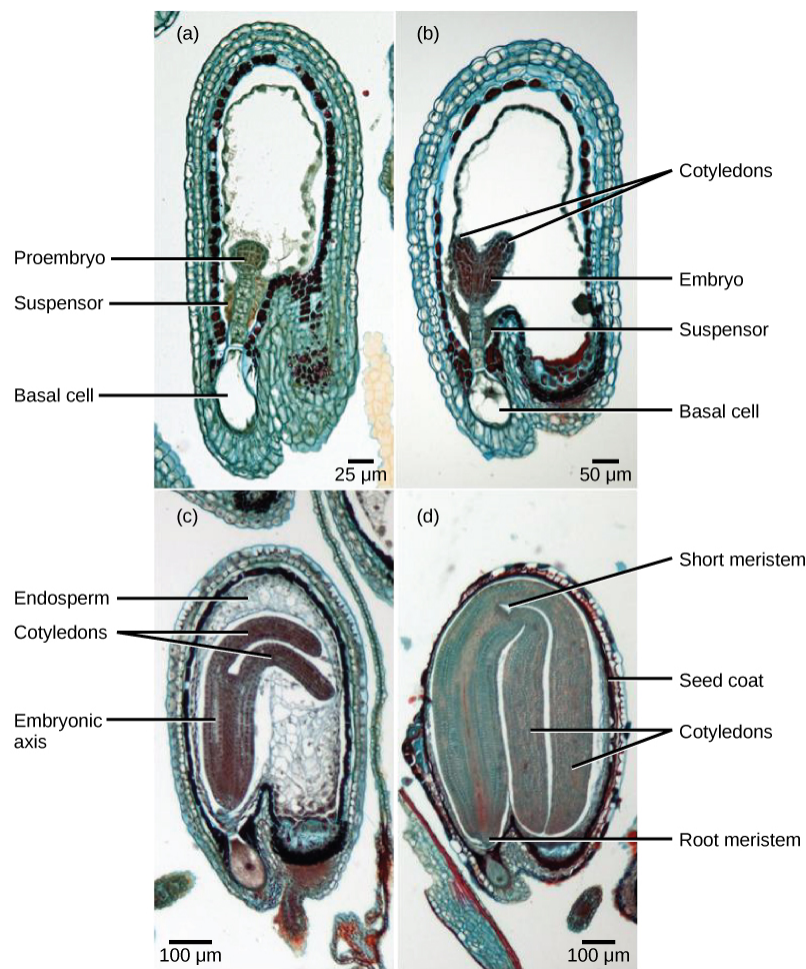
forming a triploid cell that develops into the **endosperm**. Together, these two fertilization events in angiosperms are known as **double fertilization** ([\[link\]](#)). After fertilization is complete, no other sperm can enter. The fertilized ovule forms the seed, whereas the tissues of the ovary become the fruit, usually enveloping the seed.



In angiosperms, one sperm fertilizes the egg to form the  $2n$  zygote, and the other sperm fertilizes the central cell to form the  $3n$  endosperm. This is called a double fertilization.

After fertilization, the zygote divides to form two cells: the upper cell, or terminal cell, and the lower, or basal, cell. The division of the basal cell gives rise to the **suspensor**, which eventually makes connection with the maternal tissue. The suspensor provides a route for nutrition to be transported from the mother plant to the growing embryo. The terminal cell also divides, giving rise to a globular-shaped proembryo ([\[link\]a](#)). In dicots (eudicots), the developing embryo has a heart shape, due to the presence of the two rudimentary **cotyledons** ([\[link\]b](#)). In non-endospermic dicots, such as *Capsella bursa*, the endosperm develops initially, but is then digested, and the food reserves are moved into the two cotyledons. As the embryo

and cotyledons enlarge, they run out of room inside the developing seed, and are forced to bend ([link](#)c). Ultimately, the embryo and cotyledons fill the seed ([link](#)d), and the seed is ready for dispersal. Embryonic development is suspended after some time, and growth is resumed only when the seed germinates. The developing seedling will rely on the food reserves stored in the cotyledons until the first set of leaves begin photosynthesis.



Shown are the stages of embryo development in the ovule of a shepherd's purse (*Capsella bursa*). After fertilization, the zygote divides to form an upper terminal cell and a lower basal cell. (a) In the first

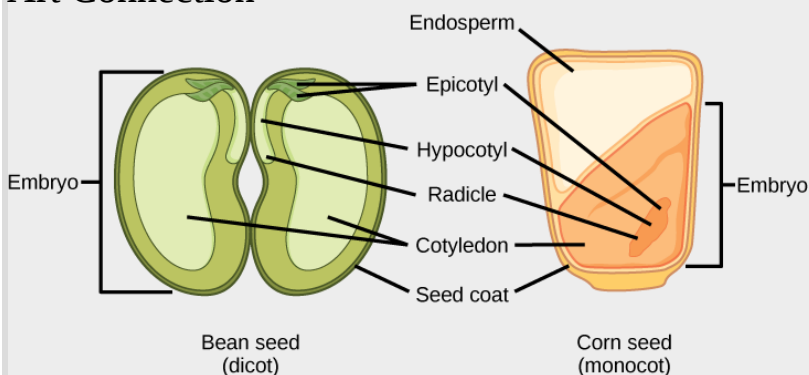
stage of development, the terminal cell divides, forming a globular pro-embryo. The basal cell also divides, giving rise to the suspensor. (b) In the second stage, the developing embryo has a heart shape due to the presence of cotyledons. (c) In the third stage, the growing embryo runs out of room and starts to bend. (d) Eventually, it completely fills the seed. (credit: modification of work by Robert R. Wise; scale-bar data from Matt Russell)

## Development of the Seed

The mature ovule develops into the seed. A typical seed contains a seed coat, cotyledons, endosperm, and a single embryo ([link](#)).

### Note:

#### Art Connection



The structures of dicot and monocot seeds are shown. Dicots (left) have two cotyledons. Monocots, such as corn (right), have one cotyledon, called the scutellum; it



channels nutrition to the growing embryo.

Both monocot and dicot embryos have a plumule that forms the leaves, a hypocotyl that forms the stem, and a radicle that forms the root. The embryonic axis comprises everything between the plumule and the radicle, not including the cotyledon(s).

What is of the following statements is true?

- a. Both monocots and dicots have an endosperm.
- b. The radicle develops into the root.
- c. The plumule is part of the epicotyl
- d. The endosperm is part of the embryo.

The storage of food reserves in angiosperm seeds differs between monocots and dicots. In monocots, such as corn and wheat, the single cotyledon is called a **scutellum**; the scutellum is connected directly to the embryo via vascular tissue (xylem and phloem). Food reserves are stored in the large endosperm. Upon germination, enzymes are secreted by the **aleurone**, a single layer of cells just inside the seed coat that surrounds the endosperm and embryo. The enzymes degrade the stored carbohydrates, proteins and lipids, the products of which are absorbed by the scutellum and transported via a vasculature strand to the developing embryo. Therefore, the scutellum can be seen to be an absorptive organ, not a storage organ.

The two cotyledons in the dicot seed also have vascular connections to the embryo. In **endospermic dicots**, the food reserves are stored in the endosperm. During germination, the two cotyledons therefore act as absorptive organs to take up the enzymatically released food reserves, much like in monocots (monocots, by definition, also have endospermic seeds). Tobacco (*Nicotiana tabacum*), tomato (*Solanum lycopersicum*), and pepper (*Capsicum annuum*) are examples of endospermic dicots. In **non-endospermic dicots**, the triploid endosperm develops normally following

double fertilization, but the endosperm food reserves are quickly remobilized and moved into the developing cotyledon for storage. The two halves of a peanut seed (*Arachis hypogaea*) and the split peas (*Pisum sativum*) of split pea soup are individual cotyledons loaded with food reserves.

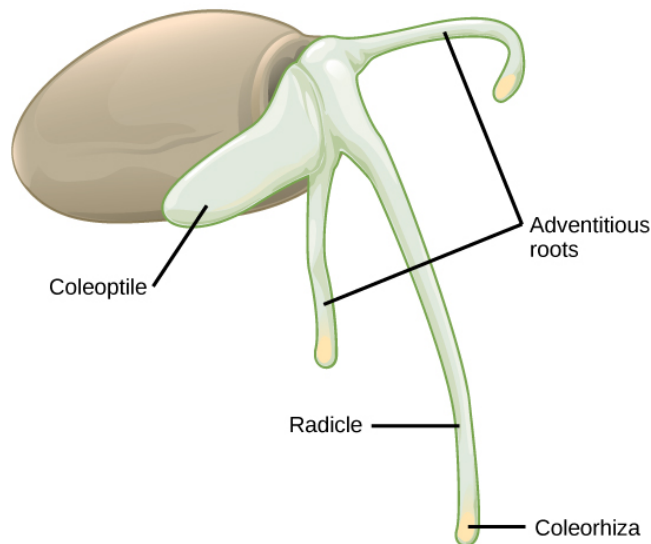
The seed, along with the ovule, is protected by a seed coat that is formed from the integuments of the ovule sac. In dicots, the seed coat is further divided into an outer coat known as the **testa** and inner coat known as the **tegmen**.

The embryonic axis consists of three parts: the plumule, the radicle, and the hypocotyl. The portion of the embryo between the cotyledon attachment point and the radicle is known as the **hypocotyl** (hypocotyl means “below the cotyledons”). The embryonic axis terminates in a **radicle** (the embryonic root), which is the region from which the root will develop. In dicots, the hypocotyls extend above ground, giving rise to the stem of the plant. In monocots, the hypocotyl does not show above ground because monocots do not exhibit stem elongation. The part of the embryonic axis that projects above the cotyledons is known as the **epicotyl**. The **plumule** is composed of the epicotyl, young leaves, and the shoot apical meristem.

Upon germination in dicot seeds, the epicotyl is shaped like a hook with the plumule pointing downwards. This shape is called the plumule hook, and it persists as long as germination proceeds in the dark. Therefore, as the epicotyl pushes through the tough and abrasive soil, the plumule is protected from damage. Upon exposure to light, the hypocotyl hook straightens out, the young foliage leaves face the sun and expand, and the epicotyl continues to elongate. During this time, the radicle is also growing and producing the primary root. As it grows downward to form the tap root, lateral roots branch off to all sides, producing the typical dicot tap root system.

In monocot seeds ([link](#)), the testa and tegmen of the seed coat are fused. As the seed germinates, the primary root emerges, protected by the root-tip covering: the **coleorhiza**. Next, the primary shoot emerges, protected by the **coleoptile**: the covering of the shoot tip. Upon exposure to light (i.e. when the plumule has exited the soil and the protective coleoptile is no longer

needed), elongation of the coleoptile ceases and the leaves expand and unfold. At the other end of the embryonic axis, the primary root soon dies, while other, adventitious roots (roots that do not arise from the usual place – i.e. the root) emerge from the base of the stem. This gives the monocot a fibrous root system.



As this monocot grass seed germinates, the primary root, or radicle, emerges first, followed by the primary shoot, or coleoptile, and the adventitious roots.

## Seed Germination

Many mature seeds enter a period of inactivity, or extremely low metabolic activity: a process known as **dormancy**, which may last for months, years or even centuries. Dormancy helps keep seeds viable during unfavorable conditions. Upon a return to favorable conditions, seed germination takes place. Favorable conditions could be as diverse as moisture, light, cold, fire,

or chemical treatments. After heavy rains, many new seedlings emerge. Forest fires also lead to the emergence of new seedlings. Some seeds require **vernalization** (cold treatment) before they can germinate. This guarantees that seeds produced by plants in temperate climates will not germinate until the spring. Plants growing in hot climates may have seeds that need a heat treatment in order to germinate, to avoid germination in the hot, dry summers. In many seeds, the presence of a thick seed coat retards the ability to germinate. **Scarification**, which includes mechanical or chemical processes to soften the seed coat, is often employed before germination. Presoaking in hot water, or passing through an acid environment, such as an animal's digestive tract, may also be employed.

Depending on seed size, the time taken for a seedling to emerge may vary. Species with large seeds have enough food reserves to germinate deep below ground, and still extend their epicotyl all the way to the soil surface. Seeds of small-seeded species usually require light as a germination cue. This ensures the seeds only germinate at or near the soil surface (where the light is greatest). If they were to germinate too far underneath the surface, the developing seedling would not have enough food reserves to reach the sunlight.

## **Development of Fruit and Fruit Types**

After fertilization, the ovary of the flower usually develops into the fruit. Fruits are usually associated with having a sweet taste; however, not all fruits are sweet. Botanically, the term “fruit” is used for a ripened ovary. In most cases, flowers in which fertilization has taken place will develop into fruits, and flowers in which fertilization has not taken place will not. Some fruits develop from the ovary and are known as true fruits, whereas others develop from other parts of the female gametophyte and are known as accessory fruits. The fruit encloses the seeds and the developing embryo, thereby providing it with protection. Fruits are of many types, depending on their origin and texture. The sweet tissue of the blackberry, the red flesh of the tomato, the shell of the peanut, and the hull of corn (the tough, thin part that gets stuck in your teeth when you eat popcorn) are all fruits. As the fruit matures, the seeds also mature.

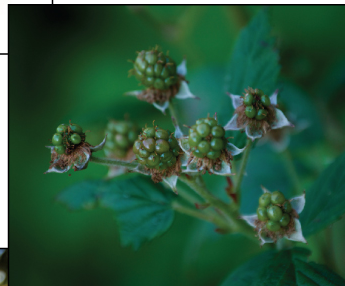
Fruits may be classified as simple, aggregate, multiple, or accessory, depending on their origin ([link](#)). If the fruit develops from a single carpel or fused carpels of a single ovary, it is known as a **simple fruit**, as seen in nuts and beans. An **aggregate fruit** is one that develops from more than one carpel, but all are in the same flower: the mature carpels fuse together to form the entire fruit, as seen in the raspberry. **Multiple fruit** develops from an inflorescence or a cluster of flowers. An example is the pineapple, where the flowers fuse together to form the fruit. **Accessory fruits** (sometimes called false fruits) are not derived from the ovary, but from another part of the flower, such as the receptacle (strawberry) or the hypanthium (apples and pears).



Simple fruit



Accessory fruit



Aggregate fruit



Multiple fruit

There are four main types of fruits. Simple fruits, such as these nuts, are derived from a

single ovary. Aggregate fruits, like raspberries, form from many carpels that fuse together. Multiple fruits, such as pineapple, form from a cluster of flowers called an inflorescence. Accessory fruit, like the apple, are formed from a part of the plant other than the ovary. (credit "nuts": modification of work by Petr Kratochvil; credit "raspberries": modification of work by Cory Zanker; credit "pineapple": modification of work by Howie Le; credit "apple": modification of work by Paolo Neo)

Fruits generally have three parts: the **exocarp** (the outermost skin or covering), the **mesocarp** (middle part of the fruit), and the **endocarp** (the inner part of the fruit). Together, all three are known as the **pericarp**. The mesocarp is usually the fleshy, edible part of the fruit; however, in some fruits, such as the almond, the endocarp is the edible part. In many fruits, two or all three of the layers are fused, and are indistinguishable at maturity. Fruits can be dry or fleshy. Furthermore, fruits can be divided into dehiscent or indehiscent types. Dehiscent fruits, such as peas, readily release their seeds, while indehiscent fruits, like peaches, rely on decay to release their seeds.

## Fruit and Seed Dispersal

The fruit has a single purpose: seed dispersal. Seeds contained within fruits need to be dispersed far from the mother plant, so they may find favorable and less competitive conditions in which to germinate and grow.

Some fruit have built-in mechanisms so they can disperse by themselves, whereas others require the help of agents like wind, water, and animals ([link](#)). Modifications in seed structure, composition, and size help in dispersal. Wind-dispersed fruit are lightweight and may have wing-like

appendages that allow them to be carried by the wind. Some have a parachute-like structure to keep them afloat. Some fruits—for example, the dandelion—have hairy, weightless structures that are suited to dispersal by wind.

Seeds dispersed by water are contained in light and buoyant fruit, giving them the ability to float. Coconuts are well known for their ability to float on water to reach land where they can germinate. Similarly, willow and silver birches produce lightweight fruit that can float on water.

Animals and birds eat fruits, and the seeds that are not digested are excreted in their droppings some distance away. Some animals, like squirrels, bury seed-containing fruits for later use; if the squirrel does not find its stash of fruit, and if conditions are favorable, the seeds germinate. Some fruits, like the cocklebur, have hooks or sticky structures that stick to an animal's coat and are then transported to another place. Humans also play a big role in dispersing seeds when they carry fruits to new places and throw away the inedible part that contains the seeds.

All of the above mechanisms allow for seeds to be dispersed through space, much like an animal's offspring can move to a new location. Seed dormancy, which was described earlier, allows plants to disperse their progeny through time: something animals cannot do. Dormant seeds can wait months, years, or even decades for the proper conditions for germination and propagation of the species.



Fruits and seeds are dispersed by various means. (a) Dandelion seeds are dispersed by wind, the (b) coconut seed is dispersed by water, and the (c) acorn is dispersed by

animals that cache and then forget it.  
(credit a: modification of work by  
"Rosendahl"/Flickr; credit b: modification  
of work by Shine Oa; credit c:  
modification of work by Paolo Neo)

## Section Summary

For fertilization to occur in angiosperms, pollen has to be transferred to the stigma of a flower: a process known as pollination. Gymnosperm pollination involves the transfer of pollen from a male cone to a female cone. When the pollen of the flower is transferred to the stigma of the same flower, it is called self-pollination. Cross-pollination occurs when pollen is transferred from one flower to another flower on the same plant, or another plant. Cross-pollination requires pollinating agents such as water, wind, or animals, and increases genetic diversity. After the pollen lands on the stigma, the tube cell gives rise to the pollen tube, through which the generative nucleus migrates. The pollen tube gains entry through the micropyle on the ovule sac. The generative cell divides to form two sperm cells: one fuses with the egg to form the diploid zygote, and the other fuses with the polar nuclei to form the endosperm, which is triploid in nature. This is known as double fertilization. After fertilization, the zygote divides to form the embryo and the fertilized ovule forms the seed. The walls of the ovary form the fruit in which the seeds develop. The seed, when mature, will germinate under favorable conditions and give rise to the diploid sporophyte.

## Art Connections

### Exercise:

**Problem:** [\[link\]](#) What is the function of the cotyledon?

- a. It develops into the root.



- b. It provides nutrition for the embryo.
- c. It forms the embryo.
- d. It protects the embryo.

---

**Solution:**

[\[link\]](#) B

## Review Questions

**Exercise:**

**Problem:** After double fertilization, a zygote and \_\_\_\_\_ form.

- a. an ovule
- b. endosperm
- c. a cotyledon
- d. a suspensor

---

**Solution:**

B

**Exercise:**

**Problem:** The fertilized ovule gives rise to the \_\_\_\_\_.

- a. fruit
- b. seed
- c. endosperm
- d. embryo

---

**Solution:**

B

**Exercise:**

**Problem:**

What is the term for a fruit that develops from tissues other than the ovary?

- a. simple fruit
- b. aggregate fruit
- c. multiple fruit
- d. accessory fruit

---

**Solution:**

D

**Exercise:**

**Problem:** The \_\_\_\_\_ is the outermost covering of a fruit.

- a. endocarp
- b. pericarp
- c. exocarp
- d. mesocarp

---

**Solution:**

C

**Free Response**

**Exercise:**

**Problem:**

Why do some seeds undergo a period of dormancy, and how do they break dormancy?

---

**Solution:**

Many seeds enter a period of inactivity or extremely low metabolic activity, a process known as dormancy. Dormancy allows seeds to tide over unfavorable conditions and germinate on return to favorable conditions. Favorable conditions could be as diverse as moisture, light, cold, fire, or chemical treatments. After heavy rains, many new seedlings emerge. Forest fires also lead to the emergence of new seedlings.

**Exercise:**

**Problem:** Discuss some ways in which fruit seeds are dispersed.

---

**Solution:**

Some fruits have built-in mechanisms that allow them to disperse seeds by themselves, but others require the assistance of agents like wind, water, and animals. Fruit that are dispersed by the wind are light in weight and often have wing-like appendages that allow them to be carried by the wind; other have structures resembling a parachute that keep them afloat in the wind. Some fruits, such as those of dandelions, have hairy, weightless structures that allow them to float in the wind. Fruits dispersed by water are light and buoyant, giving them the ability to float; coconuts are one example. Animals and birds eat fruits and disperse their seeds by leaving droppings at distant locations. Other animals bury fruit that may later germinate. Some fruits stick to animals' bodies and are carried to new locations. People also contribute to seed dispersal when they carry fruits to new places.

**Glossary**

accessory fruit

fruit derived from tissues other than the ovary

aggregate fruit

fruit that develops from multiple carpels in the same flower

aleurone

single layer of cells just inside the seed coat that secretes enzymes upon germination

coleoptile

covering of the shoot tip, found in germinating monocot seeds

coleorhiza

covering of the root tip, found in germinating monocot seeds

cotyledon

fleshy part of seed that provides nutrition to the seed

cross-pollination

transfer of pollen from the anther of one flower to the stigma of a different flower

dormancy

period of no growth and very slow metabolic processes

double fertilization

two fertilization events in angiosperms; one sperm fuses with the egg, forming the zygote, whereas the other sperm fuses with the polar nuclei, forming endosperm

endocarp

innermost part of fruit

endosperm

triploid structure resulting from fusion of a sperm with polar nuclei, which serves as a nutritive tissue for embryo

endospermic dicot

dicot that stores food reserves in the endosperm

exocarp

outermost covering of a fruit

epicotyl

embryonic shoot above the cotyledons

gravitropism

response of a plant growth in the same direction as gravity

hypocotyl

embryonic axis above the cotyledons

mesocarp

middle part of a fruit

multiple fruit

fruit that develops from multiple flowers on an inflorescence

nectar guide

pigment pattern on a flower that guides an insect to the nectaries

non-endospermic dicot

dicot that stores food reserves in the developing cotyledon

pericarp

collective term describing the exocarp, mesocarp, and endocarp; the structure that encloses the seed and is a part of the fruit

plumule

shoot that develops from the germinating seed

pollination

transfer of pollen to the stigma

radicle

original root that develops from the germinating seed

scarification

mechanical or chemical processes to soften the seed coat

scutellum

type of cotyledon found in monocots, as in grass seeds

self-pollination

transfer of pollen from the anther to the stigma of same flower

simple fruit

fruit that develops from a single carpel or fused carpels

suspensor

part of the growing embryo that makes connection with the maternal tissues

tegmen

inner layer of the seed coat

testa

outer layer of the seed coat

vernalization

exposure to cold required by some seeds before they can germinate

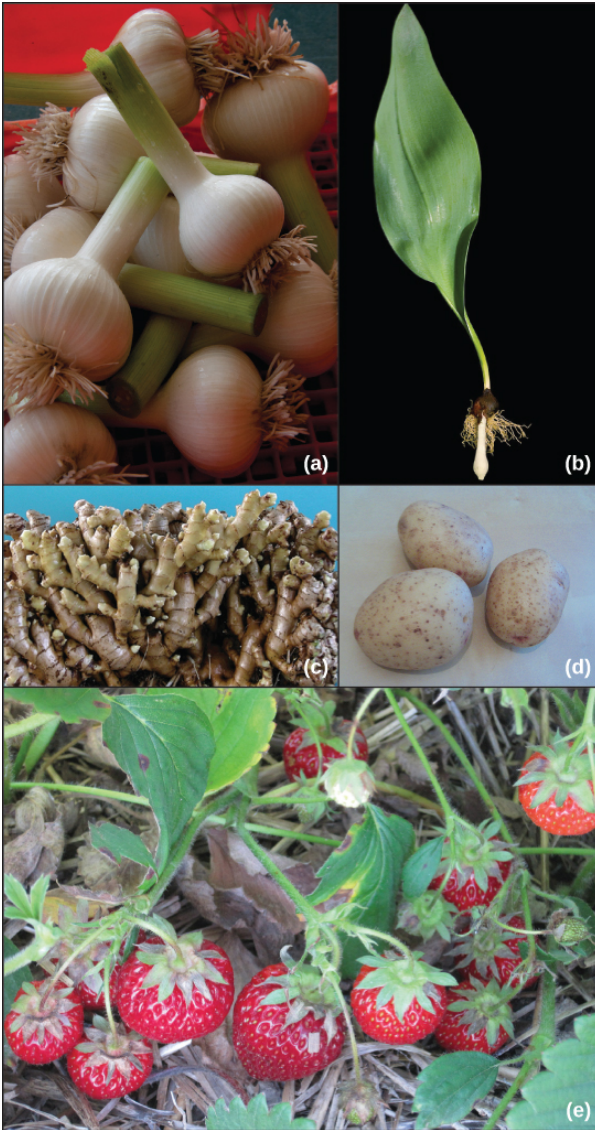
## Asexual Reproduction

By the end of this section, you will be able to:

- Compare the mechanisms and methods of natural and artificial asexual reproduction
- Describe the advantages and disadvantages of natural and artificial asexual reproduction
- Discuss plant life spans

Many plants are able to propagate themselves using asexual reproduction. This method does not require the investment required to produce a flower, attract pollinators, or find a means of seed dispersal. Asexual reproduction produces plants that are genetically identical to the parent plant because no mixing of male and female gametes takes place. Traditionally, these plants survive well under stable environmental conditions when compared with plants produced from sexual reproduction because they carry genes identical to those of their parents.

Many different types of roots exhibit asexual reproduction [\[link\]](#). The corm is used by gladiolus and garlic. Bulbs, such as a scaly bulb in lilies and a tunicate bulb in daffodils, are other common examples. A potato is a stem tuber, while parsnip propagates from a taproot. Ginger and iris produce rhizomes, while ivy uses an adventitious root (a root arising from a plant part other than the main or primary root), and the strawberry plant has a stolon, which is also called a runner.



Different types of stems allow for asexual reproduction. (a) The corm of a garlic plant looks similar to (b) a tulip bulb, but the corm is solid tissue, while the bulb consists of layers of modified leaves that surround an underground stem. Both corms and bulbs can self-propagate, giving rise to new plants. (c) Ginger forms masses



of stems called rhizomes that can give rise to multiple plants.

(d) Potato plants form fleshy stem tubers. Each eye in the stem tuber can give rise to a new plant. (e) Strawberry plants form stolons: stems that grow at the soil surface or just below ground and can give rise to new plants. (credit a: modification of work by Dwight Sipler; credit c: modification of work by Albert Cahalan, USDA ARS; credit d: modification of work by Richard North; credit e: modification of work by Julie Magro)

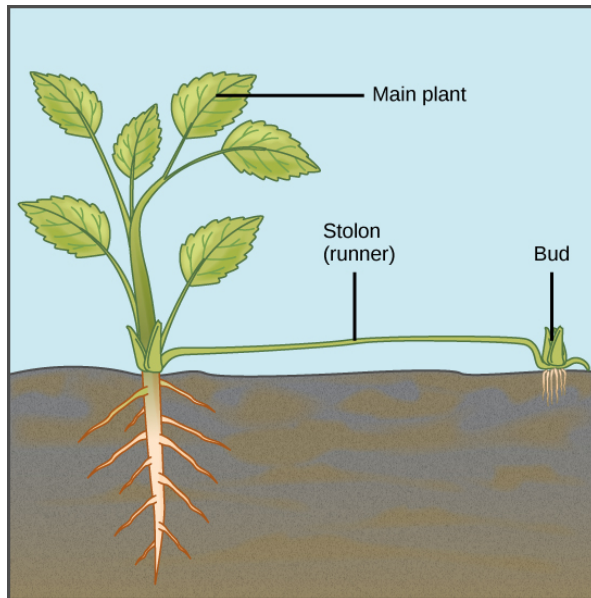
Some plants can produce seeds without fertilization. Either the ovule or part of the ovary, which is diploid in nature, gives rise to a new seed. This method of reproduction is known as **apomixis**.

An advantage of asexual reproduction is that the resulting plant will reach maturity faster. Since the new plant is arising from an adult plant or plant parts, it will also be sturdier than a seedling. Asexual reproduction can take place by natural or artificial (assisted by humans) means.

## Natural Methods of Asexual Reproduction

Natural methods of asexual reproduction include strategies that plants have developed to self-propagate. Many plants—like ginger, onion, gladioli, and dahlia—continue to grow from buds that are present on the surface of the stem. In some plants, such as the sweet potato, adventitious roots or runners can give rise to new plants [\[link\]](#). In *Bryophyllum* and kalanchoe, the leaves have small buds on their margins. When these are detached from the plant,

they grow into independent plants; or, they may start growing into independent plants if the leaf touches the soil. Some plants can be propagated through cuttings alone.



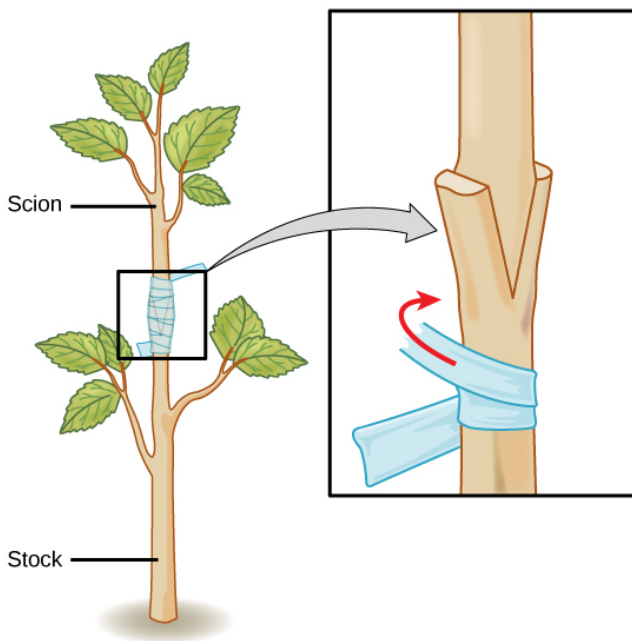
A stolon, or runner, is a stem that runs along the ground. At the nodes, it forms adventitious roots and buds that grow into a new plant.

## **Artificial Methods of Asexual Reproduction**

These methods are frequently employed to give rise to new, and sometimes novel, plants. They include grafting, cutting, layering, and micropropagation.

### **Grafting**

Grafting has long been used to produce novel varieties of roses, citrus species, and other plants. In **grafting**, two plant species are used; part of the stem of the desirable plant is grafted onto a rooted plant called the stock. The part that is grafted or attached is called the **scion**. Both are cut at an oblique angle (any angle other than a right angle), placed in close contact with each other, and are then held together [\[link\]](#). Matching up these two surfaces as closely as possible is extremely important because these will be holding the plant together. The vascular systems of the two plants grow and fuse, forming a graft. After a period of time, the scion starts producing shoots, and eventually starts bearing flowers and fruits. Grafting is widely used in viticulture (grape growing) and the citrus industry. Scions capable of producing a particular fruit variety are grafted onto root stock with specific resistance to disease.



Grafting is an artificial method of asexual reproduction used to produce plants combining favorable stem characteristics with favorable root characteristics. The stem of the plant to be grafted is

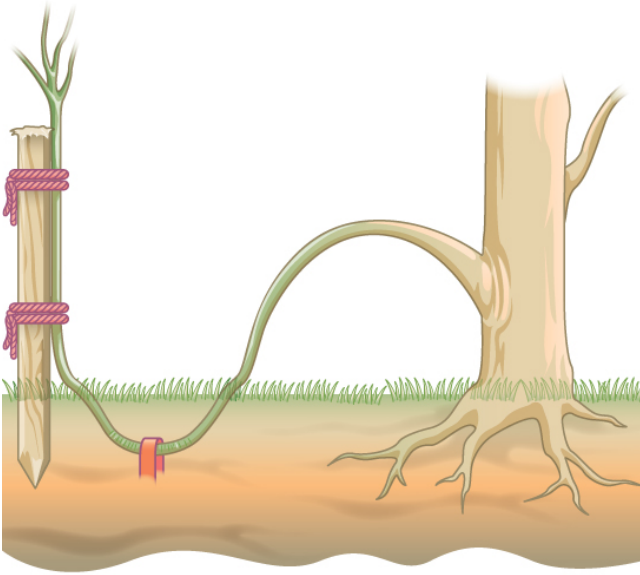
known as the scion, and the root is called the stock.

## **Cutting**

Plants such as coleus and money plant are propagated through stem **cuttings**, where a portion of the stem containing nodes and internodes is placed in moist soil and allowed to root. In some species, stems can start producing a root even when placed only in water. For example, leaves of the African violet will root if kept in water undisturbed for several weeks.

## **Layering**

**Layering** is a method in which a stem attached to the plant is bent and covered with soil. Young stems that can be bent easily without any injury are preferred. Jasmine and bougainvillea (paper flower) can be propagated this way [\[link\]](#). In some plants, a modified form of layering known as air layering is employed. A portion of the bark or outermost covering of the stem is removed and covered with moss, which is then taped. Some gardeners also apply rooting hormone. After some time, roots will appear, and this portion of the plant can be removed and transplanted into a separate pot.



In layering, a part of the stem is buried so that it forms a new plant.

(credit: modification of work by Pearson Scott Foresman, donated to the Wikimedia Foundation)

## Micropropagation

**Micropropagation** (also called plant tissue culture) is a method of propagating a large number of plants from a single plant in a short time under laboratory conditions [\[link\]](#). This method allows propagation of rare, endangered species that may be difficult to grow under natural conditions, are economically important, or are in demand as disease-free plants.



Micropropagation is used to propagate plants in sterile conditions. (credit: Nikhilesh Sanyal)

To start plant tissue culture, a part of the plant such as a stem, leaf, embryo, anther, or seed can be used. The plant material is thoroughly sterilized using a combination of chemical treatments standardized for that species. Under sterile conditions, the plant material is placed on a plant tissue culture medium that contains all the minerals, vitamins, and hormones required by the plant. The plant part often gives rise to an undifferentiated mass known as callus, from which individual plantlets begin to grow after a period of time. These can be separated and are first grown under greenhouse conditions before they are moved to field conditions.

## **Plant Life Spans**

The length of time from the beginning of development to the death of a plant is called its life span. The life cycle, on the other hand, is the sequence of stages a plant goes through from seed germination to seed production of the mature plant. Some plants, such as annuals, only need a few weeks to grow, produce seeds and die. Other plants, such as the bristlecone pine, live for thousands of years. Some bristlecone pines have a documented age of 4,500 years [\[link\]](#). Even as some parts of a plant, such as regions containing meristematic tissue—the area of active plant growth consisting of undifferentiated cells capable of cell division—continue to grow, some parts undergo programmed cell death (apoptosis). The cork found on stems, and the water-conducting tissue of the xylem, for example, are composed of dead cells.



The bristlecone pine, shown here in the Ancient Bristlecone Pine Forest in the White Mountains of eastern California, has been known to live for 4,500 years. (credit: Rick Goldwaser)

Plant species that complete their lifecycle in one season are known as annuals, an example of which is *Arabidopsis*, or mouse-ear cress. Biennials such as carrots complete their lifecycle in two seasons. In a biennial's first

season, the plant has a vegetative phase, whereas in the next season, it completes its reproductive phase. Commercial growers harvest the carrot roots after the first year of growth, and do not allow the plants to flower. Perennials, such as the magnolia, complete their lifecycle in two years or more.

In another classification based on flowering frequency, **monocarpic** plants flower only once in their lifetime; examples include bamboo and yucca. During the vegetative period of their life cycle (which may be as long as 120 years in some bamboo species), these plants may reproduce asexually and accumulate a great deal of food material that will be required during their once-in-a-lifetime flowering and setting of seed after fertilization. Soon after flowering, these plants die. **Polycarpic** plants form flowers many times during their lifetime. Fruit trees, such as apple and orange trees, are polycarpic; they flower every year. Other polycarpic species, such as perennials, flower several times during their life span, but not each year. By this means, the plant does not require all its nutrients to be channelled towards flowering each year.

As is the case with all living organisms, genetics and environmental conditions have a role to play in determining how long a plant will live. Susceptibility to disease, changing environmental conditions, drought, cold, and competition for nutrients are some of the factors that determine the survival of a plant. Plants continue to grow, despite the presence of dead tissue such as cork. Individual parts of plants, such as flowers and leaves, have different rates of survival. In many trees, the older leaves turn yellow and eventually fall from the tree. Leaf fall is triggered by factors such as a decrease in photosynthetic efficiency, due to shading by upper leaves, or oxidative damage incurred as a result of photosynthetic reactions. The components of the part to be shed are recycled by the plant for use in other processes, such as development of seed and storage. This process is known as nutrient recycling.

The aging of a plant and all the associated processes is known as **senescence**, which is marked by several complex biochemical changes. One of the characteristics of senescence is the breakdown of chloroplasts, which is characterized by the yellowing of leaves. The chloroplasts contain



components of photosynthetic machinery such as membranes and proteins. Chloroplasts also contain DNA. The proteins, lipids, and nucleic acids are broken down by specific enzymes into smaller molecules and salvaged by the plant to support the growth of other plant tissues.

The complex pathways of nutrient recycling within a plant are not well understood. Hormones are known to play a role in senescence. Applications of cytokinins and ethylene delay or prevent senescence; in contrast, abscissic acid causes premature onset of senescence.

## **Sections Summary**

Many plants reproduce asexually as well as sexually. In asexual reproduction, part of the parent plant is used to generate a new plant. Grafting, layering, and micropropagation are some methods used for artificial asexual reproduction. The new plant is genetically identical to the parent plant from which the stock has been taken. Asexually reproducing plants thrive well in stable environments.

Plants have different life spans, dependent on species, genotype, and environmental conditions. Parts of the plant, such as regions containing meristematic tissue, continue to grow, while other parts experience programmed cell death. Leaves that are no longer photosynthetically active are shed from the plant as part of senescence, and the nutrients from these leaves are recycled by the plant. Other factors, including the presence of hormones, are known to play a role in delaying senescence.

## **Review Questions**

### **Exercise:**

#### **Problem:**

\_\_\_\_\_ is a useful method of asexual reproduction for propagating hard-to-root plants.

- a. grafting

- b. layering
- c. cuttings
- d. budding

---

**Solution:**

A

**Exercise:**

**Problem:**

Which of the following is an advantage of asexual reproduction?

- a. Cuttings taken from an adult plant show increased resistance to diseases.
- b. Grafted plants can more successfully endure drought.
- c. When cuttings or buds are taken from an adult plant or plant parts, the resulting plant will grow into an adult faster than a seedling.
- d. Asexual reproduction takes advantage of a more diverse gene pool.

---

**Solution:**

C

**Exercise:**

**Problem:**

Plants that flower once in their lifetime are known as \_\_\_\_\_.

- a. monoecious
  - b. dioecious
  - c. polycarpic
  - d. monocarpic
-

**Solution:**

D

**Exercise:**

**Problem:**

Plant species that complete their lifecycle in one season are known as \_\_\_\_\_.

- a. biennials
- b. perennials
- c. annuals
- d. polycarpic

---

**Solution:**

C

**Free Response**

**Exercise:**

**Problem:**

What are some advantages of asexual reproduction in plants?

---

**Solution:**

Asexual reproduction does not require the expenditure of the plant's resources and energy that would be involved in producing a flower, attracting pollinators, or dispersing seeds. Asexual reproduction results in plants that are genetically identical to the parent plant, since there is no mixing of male and female gametes, resulting in better survival. The cuttings or buds taken from an adult plant produce progeny that mature faster and are sturdier than a seedling grown from a seed.

**Exercise:****Problem:**

Describe natural and artificial methods of asexual reproduction in plants.

---

**Solution:**

Asexual reproduction in plants can take place by natural methods or artificial methods. Natural methods include strategies used by the plant to propagate itself. Artificial methods include grafting, cutting, layering, and micropropagation.

**Exercise:**

**Problem:** Discuss the life cycles of various plants.

---

**Solution:**

Plant species that complete their life cycle in one season are known as annuals. Biennials complete their life cycle in two seasons. In the first season, the plant has a vegetative phase, whereas in the next season, it completes its reproductive phase. Perennials, such as the magnolia, complete their life cycle in two years or more.

**Exercise:****Problem:**

How are plants classified on the basis of flowering frequency?

---

**Solution:**

Monocarpic plants flower only once during their lifetime. During the vegetative period of their lifecycle, these plants accumulate a great deal of food material that will be required during their once-in-a-lifetime flowering and setting of seed after fertilization. Soon after flowering, these plants die. Polycarpic plants flower several times

during their life span; therefore, not all nutrients are channelled towards flowering.

## **Glossary**

apomixis

process by which seeds are produced without fertilization of sperm and egg

cutting

method of asexual reproduction where a portion of the stem contains nodes and internodes is placed in moist soil and allowed to root

grafting

method of asexual reproduction where the stem from one plant species is spliced to a different plant

layering

method of propagating plants by bending a stem under the soil

micropropagation

propagation of desirable plants from a plant part; carried out in a laboratory

monocarpic

plants that flower once in their lifetime

polycarpic

plants that flower several times in their lifetime

scion

the part of a plant that is grafted onto the root stock of another plant

senescence

process that describes aging in plant tissues

## Introduction

class="introduction"

The (a) deer  
tick carries  
the  
bacterium  
that  
produces  
Lyme  
disease in  
humans,  
often  
evident in

(b) a  
symptomatic bull's eye  
rash. The (c)  
white-footed  
mouse is  
one well-  
known host  
to deer ticks  
carrying the  
Lyme  
disease  
bacterium.

(credit a:  
modification  
of work by  
Scott Bauer,  
USDA  
ARS; credit  
b:  
modification  
of work by  
James

Gathany,  
CDC; credit  
c:  
modification  
of work by  
Rob Ireton)



(a)



(b)



(c)

Why study ecology? Perhaps you are interested in learning about the natural world and how living things have adapted to the physical conditions of their environment. Or, perhaps you're a future physician seeking to understand the connection between human health and ecology.

Humans are a part of the ecological landscape, and human health is one important part of human interaction with our physical and living environment. Lyme disease, for instance, serves as one modern-day example of the connection between our health and the natural world ([link](#)). More formally known as Lyme borreliosis, Lyme disease is a bacterial infection that can be transmitted to humans when they are bitten by the deer tick (*Ixodes scapularis*), which is the primary vector for this disease. However, not all deer ticks carry the bacteria that will cause Lyme disease in humans, and *I. scapularis* can have other hosts besides deer. In fact, it turns out that the probability of infection depends on the type of host upon which the tick develops: a higher proportion of ticks that live on

white-footed mice carry the bacterium than do ticks that live on deer. Knowledge about the environments and population densities in which the host species is abundant would help a physician or an epidemiologist better understand how Lyme disease is transmitted and how its incidence could be reduced.



## The Scope of Ecology

By the end of this section, you will be able to:

- Define ecology and the four levels of ecological research
- Describe examples of the ways in which ecology requires the integration of different scientific disciplines
- Distinguish between abiotic and biotic components of the environment
- Recognize the relationship between abiotic and biotic components of the environment

**Ecology** is the study of the interactions of living organisms with their environment. One core goal of ecology is to understand the distribution and abundance of living things in the physical environment. Attainment of this goal requires the integration of scientific disciplines inside and outside of biology, such as biochemistry, physiology, evolution, biodiversity, molecular biology, geology, and climatology. Some ecological research also applies aspects of chemistry and physics, and it frequently uses mathematical models.

### Note:

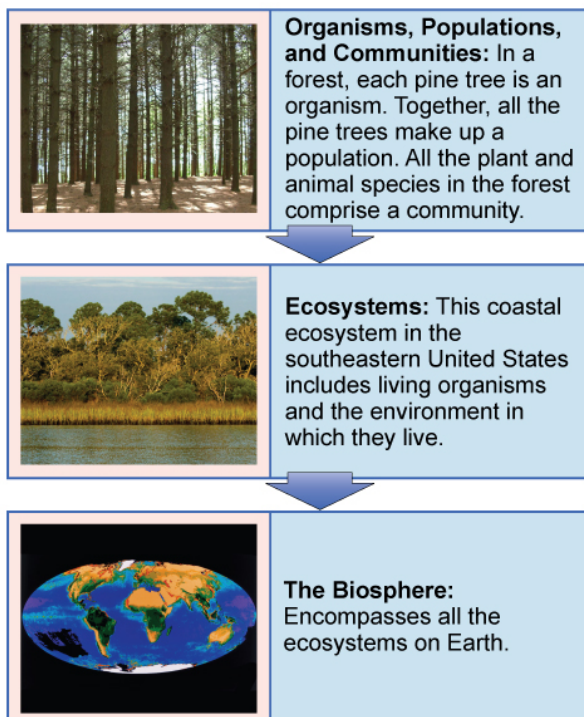
#### Link to Learning



Climate change can alter where organisms live, which can sometimes directly affect human health. Watch the PBS video [“Feeling the Effects of Climate Change”](#) in which researchers discover a pathogenic organism living far outside of its normal range.

## Levels of Ecological Study

When a discipline such as biology is studied, it is often helpful to subdivide it into smaller, related areas. For instance, cell biologists interested in cell signaling need to understand the chemistry of the signal molecules (which are usually proteins) as well as the result of cell signaling. Ecologists interested in the factors that influence the survival of an endangered species might use mathematical models to predict how current conservation efforts affect endangered organisms. To produce a sound set of management options, a conservation biologist needs to collect accurate data, including current population size, factors affecting reproduction (like physiology and behavior), habitat requirements (such as plants and soils), and potential human influences on the endangered population and its habitat (which might be derived through studies in sociology and urban ecology). Within the discipline of ecology, researchers work at four specific levels, sometimes discretely and sometimes with overlap: organism, population, community, and ecosystem ([link](#)).



Ecologists study within several  
biological levels of

organization. (credit  
“organisms”: modification of  
work by "Crystl"/Flickr; credit  
“ecosystems”: modification of  
work by Tom Carlisle, US Fish  
and Wildlife Service  
Headquarters; credit  
“biosphere”: NASA)

## Organismal Ecology

Researchers studying ecology at the organismal level are interested in the adaptations that enable individuals to live in specific habitats. These adaptations can be morphological, physiological, and behavioral. For instance, the Karner blue butterfly (*Lycaeides melissa samuelis*) ([link](#)) is considered a specialist because the females preferentially oviposit (that is, lay eggs) on wild lupine. This preferential adaptation means that the Karner blue butterfly is highly dependent on the presence of wild lupine plants for its continued survival.



The Karner blue butterfly  
(*Lycaeides melissa samuelis*)  
is a rare butterfly that lives

only in open areas with few trees or shrubs, such as pine barrens and oak savannas. It can only lay its eggs on lupine plants. (credit: modification of work by J & K Hollingsworth, USFWS)

After hatching, the larval caterpillars emerge and spend four to six weeks feeding solely on wild lupine ([link](#)). The caterpillars pupate (undergo metamorphosis) and emerge as butterflies after about four weeks. The adult butterflies feed on the nectar of flowers of wild lupine and other plant species. A researcher interested in studying Karner blue butterflies at the organismal level might, in addition to asking questions about egg laying, ask questions about the butterflies' preferred temperature (a physiological question) or the behavior of the caterpillars when they are at different larval stages (a behavioral question).



The wild lupine  
(*Lupinus  
perennis*) is the  
host plant for the  
Karner blue  
butterfly.

## Population Ecology

A population is a group of interbreeding organisms that are members of the same species living in the same area at the same time. (Organisms that are all members of the same species are called **conspecifics**.) A population is identified, in part, by where it lives, and its area of population may have natural or artificial boundaries: natural boundaries might be rivers, mountains, or deserts, while examples of artificial boundaries include mowed grass, manmade structures, or roads. The study of population ecology focuses on the number of individuals in an area and how and why population size changes over time. Population ecologists are particularly interested in counting the Karner blue butterfly, for example, because it is classified as federally endangered. However, the distribution and density of this species is highly influenced by the distribution and abundance of wild lupine. Researchers might ask questions about the factors leading to the decline of wild lupine and how these affect Karner blue butterflies. For example, ecologists know that wild lupine thrives in open areas where trees and shrubs are largely absent. In natural settings, intermittent wildfires regularly remove trees and shrubs, helping to maintain the open areas that wild lupine requires. Mathematical models can be used to understand how wildfire suppression by humans has led to the decline of this important plant for the Karner blue butterfly.

## Community Ecology

A biological community consists of the different species within an area, typically a three-dimensional space, and the interactions within and among

these species. Community ecologists are interested in the processes driving these interactions and their consequences. Questions about conspecific interactions often focus on competition among members of the same species for a limited resource. Ecologists also study interactions among various species; members of different species are called **heterospecifics**. Examples of heterospecific interactions include predation, parasitism, herbivory, competition, and pollination. These interactions can have regulating effects on population sizes and can impact ecological and evolutionary processes affecting diversity.

For example, Karner blue butterfly larvae form mutualistic relationships with ants. Mutualism is a form of a long-term relationship that has coevolved between two species and from which each species benefits. For mutualism to exist between individual organisms, each species must receive some benefit from the other as a consequence of the relationship. Researchers have shown that there is an increase in the probability of survival when Karner blue butterfly larvae (caterpillars) are tended by ants. This might be because the larvae spend less time in each life stage when tended by ants, which provides an advantage for the larvae. Meanwhile, the Karner blue butterfly larvae secrete a carbohydrate-rich substance that is an important energy source for the ants. Both the Karner blue larvae and the ants benefit from their interaction.

## **Ecosystem Ecology**

Ecosystem ecology is an extension of organismal, population, and community ecology. The ecosystem is composed of all the **biotic** components (living things) in an area along with the **abiotic** components (non-living things) of that area. Some of the abiotic components include air, water, and soil. Ecosystem biologists ask questions about how nutrients and energy are stored and how they move among organisms and the surrounding atmosphere, soil, and water.

The Karner blue butterflies and the wild lupine live in an oak-pine barren habitat. This habitat is characterized by natural disturbance and nutrient-poor soils that are low in nitrogen. The availability of nutrients is an important factor in the distribution of the plants that live in this habitat.

Researchers interested in ecosystem ecology could ask questions about the importance of limited resources and the movement of resources, such as nutrients, through the biotic and abiotic portions of the ecosystem.

**Note:**

**Career Connection**

**Ecologist**

A career in ecology contributes to many facets of human society. Understanding ecological issues can help society meet the basic human needs of food, shelter, and health care. Ecologists can conduct their research in the laboratory and outside in natural environments ([\[link\]](#)). These natural environments can be as close to home as the stream running through your campus or as far away as the hydrothermal vents at the bottom of the Pacific Ocean. Ecologists manage natural resources such as white-tailed deer populations (*Odocoileus virginianus*) for hunting or aspen (*Populus* spp.) timber stands for paper production. Ecologists also work as educators who teach children and adults at various institutions including universities, high schools, museums, and nature centers. Ecologists may also work in advisory positions assisting local, state, and federal policymakers to develop laws that are ecologically sound, or they may develop those policies and legislation themselves. To become an ecologist requires an undergraduate degree, usually in a natural science. The undergraduate degree is often followed by specialized training or an advanced degree, depending on the area of ecology selected. Ecologists should also have a broad background in the physical sciences, as well as a sound foundation in mathematics and statistics.



This landscape ecologist is releasing a black-footed ferret into its native habitat as part of a study. (credit: USFWS Mountain Prairie Region, NPS)

**Note:**

Link to Learning



Visit this [site](#) to see Stephen Wing, a marine ecologist from the University of Otago, discuss the role of an ecologist and the types of issues ecologists explore.

## Section Summary

Ecology is the study of the interactions of living things with their environment. Ecologists ask questions across four levels of biological organization—organismal, population, community, and ecosystem. At the organismal level, ecologists study individual organisms and how they interact with their environments. At the population and community levels, ecologists explore, respectively, how a population of organisms changes over time and the ways in which that population interacts with other species in the community. Ecologists studying an ecosystem examine the living species (the biotic components) of the ecosystem as well as the nonliving



portions (the abiotic components), such as air, water, and soil, of the environment.

## Review Questions

### Exercise:

**Problem:** Which of the following is a biotic factor?

- a. wind
- b. disease-causing microbe
- c. temperature
- d. soil particle size

---

### Solution:

B

### Exercise:

**Problem:**

The study of nutrient cycling through the environment is an example of which of the following?

- a. organismal ecology
- b. population ecology
- c. community ecology
- d. ecosystem ecology

---

### Solution:

D

## Free Response

**Exercise:****Problem:**

Ecologists often collaborate with other researchers interested in ecological questions. Describe the levels of ecology that would be easier for collaboration because of the similarities of questions asked. What levels of ecology might be more difficult for collaboration?

---

**Solution:**

Ecologists working in organismal or population ecology might ask similar questions about how the biotic and abiotic conditions affect particular organisms and, thus, might find collaboration to be mutually beneficial. Levels of ecology such as community ecology or ecosystem ecology might pose greater challenges for collaboration because these areas are very broad and may include many different environmental components.

**Exercise:****Problem:**

The population is an important unit in ecology as well as other biological sciences. How is a population defined, and what are the strengths and weaknesses of this definition? Are there some species that at certain times or places are not in populations?

---

**Solution:**

It is beneficial to consider a population to be all of the individuals living in the same area at the same time because it allows the ecologist to identify and study all of the abiotic and biotic factors that may affect the members of the population. However, this definition of a population could be considered a drawback if it prohibits the ecologist from studying a population's individuals that may be transitory, but still influential. Some species with members that have a wide geographic range might not be considered to be a population, but could still have many of the qualities of a population.

## **Glossary**

abiotic

nonliving components of the environment

biotic

living components of the environment

conspecifics

individuals that are members of the same species

ecology

study of interaction between living things and their environment

heterospecifics

individuals that are members of different species

## Biogeography

By the end of this section, you will be able to:

- Define biogeography
- List and describe abiotic factors that affect the global distribution of plant and animal species
- Compare the impact of abiotic forces on aquatic and terrestrial environments
- Summarize the affect of abiotic factors on net primary productivity

Many forces influence the communities of living organisms present in different parts of the biosphere (all of the parts of Earth inhabited by life). The biosphere extends into the atmosphere (several kilometers above Earth) and into the depths of the oceans. Despite its apparent vastness to an individual human, the biosphere occupies only a minute space when compared to the known universe. Many abiotic forces influence where life can exist and the types of organisms found in different parts of the biosphere. The abiotic factors influence the distribution of **biomes**: large areas of land with similar climate, flora, and fauna.

## Biogeography

**Biogeography** is the study of the geographic distribution of living things and the abiotic factors that affect their distribution. Abiotic factors such as temperature and rainfall vary based mainly on latitude and elevation. As these abiotic factors change, the composition of plant and animal communities also changes. For example, if you were to begin a journey at the equator and walk north, you would notice gradual changes in plant communities. At the beginning of your journey, you would see tropical wet forests with broad-leaved evergreen trees, which are characteristic of plant communities found near the equator. As you continued to travel north, you would see these broad-leaved evergreen plants eventually give rise to seasonally dry forests with scattered trees. You would also begin to notice changes in temperature and moisture. At about 30 degrees north, these forests would give way to deserts, which are characterized by low precipitation.

Moving farther north, you would see that deserts are replaced by grasslands or prairies. Eventually, grasslands are replaced by deciduous temperate forests. These deciduous forests give way to the boreal forests found in the subarctic, the area south of the Arctic Circle. Finally, you would reach the Arctic tundra, which is found at the most northern latitudes. This trek north reveals gradual changes in both climate and the types of organisms that have adapted to environmental factors associated with ecosystems found at different latitudes. However, different ecosystems exist at the same latitude due in part to abiotic factors such as jet streams, the Gulf Stream, and ocean currents. If you were to hike up a mountain, the changes you would see in the vegetation would parallel those as you move to higher latitudes.

Ecologists who study biogeography examine patterns of species distribution. No species exists everywhere; for example, the Venus flytrap is endemic to a small area in North and South Carolina. An **endemic** species is one which is naturally found only in a specific geographic area that is usually restricted in size. Other species are generalists: species which live in a wide variety of geographic areas; the raccoon, for example, is native to most of North and Central America.

Species distribution patterns are based on biotic and abiotic factors and their influences during the very long periods of time required for species evolution; therefore, early studies of biogeography were closely linked to the emergence of evolutionary thinking in the eighteenth century. Some of the most distinctive assemblages of plants and animals occur in regions that have been physically separated for millions of years by geographic barriers. Biologists estimate that Australia, for example, has between 600,000 and 700,000 species of plants and animals. Approximately 3/4 of living plant and mammal species are endemic species found solely in Australia ([link](#)ab).



(a)



(b)

Australia is home to many endemic species. The (a) wallaby (*Wallabia bicolor*), a medium-sized member of the kangaroo family, is a pouched mammal, or marsupial. The (b) echidna (*Tachyglossus aculeatus*) is an egg-laying mammal. (credit a: modification of work by Derrick Coetzee; credit b: modification of work by Allan Whittome)

Sometimes ecologists discover unique patterns of species distribution by determining where species are *not* found. Hawaii, for example, has no native land species of reptiles or amphibians, and has only one native terrestrial mammal, the hoary bat. Most of New Guinea, as another example, lacks placental mammals.

**Note:**

Link to Learning



Check out this [video](#) to observe a platypus swimming in its natural habitat in New South Wales, Australia.

Plants can be endemic or generalists: endemic plants are found only on specific regions of the Earth, while generalists are found on many regions. Isolated land masses—such as Australia, Hawaii, and Madagascar—often have large numbers of endemic plant species. Some of these plants are endangered due to human activity. The forest gardenia (*Gardenia brighamii*), for instance, is endemic to Hawaii; only an estimated 15–20 trees are thought to exist ([link](#)).



Listed as federally endangered, the forest gardenia is a small tree with distinctive flowers. It is found only in five of the Hawaiian Islands in small

populations consisting of a few individual specimens.

(credit: Forest & Kim Starr)

## Energy Sources

Energy from the sun is captured by green plants, algae, cyanobacteria, and photosynthetic protists. These organisms convert solar energy into the chemical energy needed by all living things. Light availability can be an important force directly affecting the evolution of adaptations in photosynthesizers. For instance, plants in the understory of a temperate forest are shaded when the trees above them in the canopy completely leaf out in the late spring. Not surprisingly, understory plants have adaptations to successfully capture available light. One such adaptation is the rapid growth of spring ephemeral plants such as the spring beauty ([\[link\]](#)). These spring flowers achieve much of their growth and finish their life cycle (reproduce) early in the season before the trees in the canopy develop leaves.

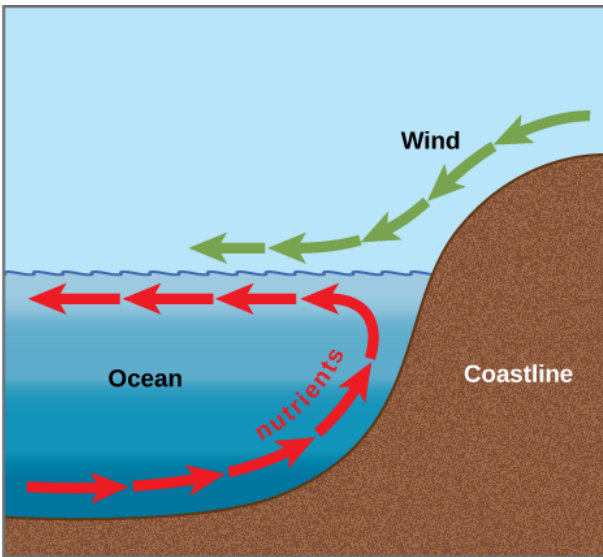




The spring beauty is an ephemeral spring plant that flowers early in the spring to avoid competing with larger forest trees for sunlight. (credit: John Beetham)

In aquatic ecosystems, the availability of light may be limited because sunlight is absorbed by water, plants, suspended particles, and resident microorganisms. Toward the bottom of a lake, pond, or ocean, there is a zone that light cannot reach. Photosynthesis cannot take place there and, as a result, a number of adaptations have evolved that enable living things to survive without light. For instance, aquatic plants have photosynthetic tissue near the surface of the water; for example, think of the broad, floating leaves of a water lily—water lilies cannot survive without light. In environments such as hydrothermal vents, some bacteria extract energy from inorganic chemicals because there is no light for photosynthesis.

The availability of nutrients in aquatic systems is also an important aspect of energy or photosynthesis. Many organisms sink to the bottom of the ocean when they die in the open water; when this occurs, the energy found in that living organism is sequestered for some time unless ocean upwelling occurs. **Ocean upwelling** is the rising of deep ocean waters that occurs when prevailing winds blow along surface waters near a coastline ([\[link\]](#)). As the wind pushes ocean waters offshore, water from the bottom of the ocean moves up to replace this water. As a result, the nutrients once contained in dead organisms become available for reuse by other living organisms.



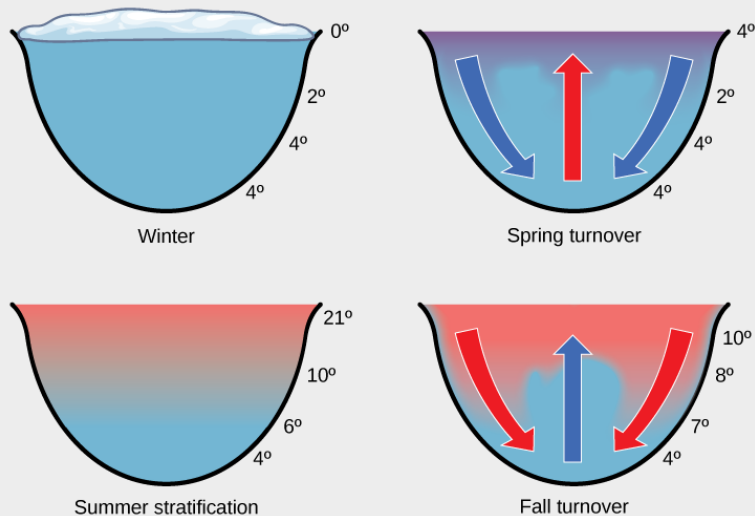
Ocean upwelling is an important process that recycles nutrients and energy in the ocean. As wind (green arrows) pushes offshore, it causes water from the ocean bottom (red arrows) to move to the surface, bringing up nutrients from the ocean depths.

In freshwater systems, the recycling of nutrients occurs in response to air temperature changes. The nutrients at the bottom of lakes are recycled twice each year: in the spring and fall turnover. The **spring and fall turnover** is a seasonal process that recycles nutrients and oxygen from the bottom of a freshwater ecosystem to the top of a body of water ([link](#)). These turnovers are caused by the formation of a **thermocline**: a layer of water with a temperature that is significantly different from that of the surrounding layers. In wintertime, the surface of lakes found in many northern regions is frozen. However, the water under the ice is slightly warmer, and the water at the bottom of the lake is warmer yet at 4 °C to 5 °C (39.2 °F to 41 °F). Water is densest at 4 °C; therefore, the deepest water is also the densest. The deepest water is oxygen poor because the decomposition of organic material at the bottom of the lake uses up available oxygen that cannot be

replaced by means of oxygen diffusion into the water due to the surface ice layer.

**Note:**

**Art Connection**



The spring and fall turnovers are important processes in freshwater lakes that act to move the nutrients and oxygen at the bottom of deep lakes to the top. Turnover occurs because water has a maximum density at 4 °C. Surface water temperature changes as the seasons progress, and denser water sinks.

How might turnover in tropical lakes differ from turnover in lakes that exist in temperate regions?

In springtime, air temperatures increase and surface ice melts. When the temperature of the surface water begins to reach 4 °C, the water becomes heavier and sinks to the bottom. The water at the bottom of the lake is then

displaced by the heavier surface water and, thus, rises to the top. As that water rises to the top, the sediments and nutrients from the lake bottom are brought along with it. During the summer months, the lake water stratifies, or forms layers, with the warmest water at the lake surface.

As air temperatures drop in the fall, the temperature of the lake water cools to 4 °C; therefore, this causes fall turnover as the heavy cold water sinks and displaces the water at the bottom. The oxygen-rich water at the surface of the lake then moves to the bottom of the lake, while the nutrients at the bottom of the lake rise to the surface ([link](#)). During the winter, the oxygen at the bottom of the lake is used by decomposers and other organisms requiring oxygen, such as fish.

## **Temperature**

Temperature affects the physiology of living things as well as the density and state of water. Temperature exerts an important influence on living things because few living things can survive at temperatures below 0 °C (32 °F) due to metabolic constraints. It is also rare for living things to survive at temperatures exceeding 45 °C (113 °F); this is a reflection of evolutionary response to typical temperatures. Enzymes are most efficient within a narrow and specific range of temperatures; enzyme degradation can occur at higher temperatures. Therefore, organisms either must maintain an internal temperature or they must inhabit an environment that will keep the body within a temperature range that supports metabolism. Some animals have adapted to enable their bodies to survive significant temperature fluctuations, such as seen in hibernation or reptilian torpor. Similarly, some bacteria are adapted to surviving in extremely hot temperatures such as geysers. Such bacteria are examples of extremophiles: organisms that thrive in extreme environments.

Temperature can limit the distribution of living things. Animals faced with temperature fluctuations may respond with adaptations, such as migration, in order to survive. Migration, the movement from one place to another, is an adaptation found in many animals, including many that inhabit seasonally cold climates. Migration solves problems related to temperature,

locating food, and finding a mate. In migration, for instance, the Arctic Tern (*Sterna paradisaea*) makes a 40,000 km (24,000 mi) round trip flight each year between its feeding grounds in the southern hemisphere and its breeding grounds in the Arctic Ocean. Monarch butterflies (*Danaus plexippus*) live in the eastern United States in the warmer months and migrate to Mexico and the southern United States in the wintertime. Some species of mammals also make migratory forays. Reindeer (*Rangifer tarandus*) travel about 5,000 km (3,100 mi) each year to find food. Amphibians and reptiles are more limited in their distribution because they lack migratory ability. Not all animals that can migrate do so: migration carries risk and comes at a high energy cost.

Some animals hibernate or estivate to survive hostile temperatures. Hibernation enables animals to survive cold conditions, and estivation allows animals to survive the hostile conditions of a hot, dry climate. Animals that hibernate or estivate enter a state known as torpor: a condition in which their metabolic rate is significantly lowered. This enables the animal to wait until its environment better supports its survival. Some amphibians, such as the wood frog (*Rana sylvatica*), have an antifreeze-like chemical in their cells, which retains the cells' integrity and prevents them from bursting.

## **Water**

Water is required by all living things because it is critical for cellular processes. Since terrestrial organisms lose water to the environment by simple diffusion, they have evolved many adaptations to retain water.

- Plants have a number of interesting features on their leaves, such as leaf hairs and a waxy cuticle, that serve to decrease the rate of water loss via transpiration.
- Freshwater organisms are surrounded by water and are constantly in danger of having water rush into their cells because of osmosis. Many adaptations of organisms living in freshwater environments have evolved to ensure that solute concentrations in their bodies remain

within appropriate levels. One such adaptation is the excretion of dilute urine.

- Marine organisms are surrounded by water with a higher solute concentration than the organism and, thus, are in danger of losing water to the environment because of osmosis. These organisms have morphological and physiological adaptations to retain water and release solutes into the environment. For example, Marine iguanas (*Amblyrhynchus cristatus*), sneeze out water vapor that is high in salt in order to maintain solute concentrations within an acceptable range while swimming in the ocean and eating marine plants.

## **Inorganic Nutrients and Soil**

Inorganic nutrients, such as nitrogen and phosphorus, are important in the distribution and the abundance of living things. Plants obtain these inorganic nutrients from the soil when water moves into the plant through the roots. Therefore, soil structure (particle size of soil components), soil pH, and soil nutrient content play an important role in the distribution of plants. Animals obtain inorganic nutrients from the food they consume. Therefore, animal distributions are related to the distribution of what they eat. In some cases, animals will follow their food resource as it moves through the environment.

## **Other Aquatic Factors**

Some abiotic factors, such as oxygen, are important in aquatic ecosystems as well as terrestrial environments. Terrestrial animals obtain oxygen from the air they breathe. Oxygen availability can be an issue for organisms living at very high elevations, however, where there are fewer molecules of oxygen in the air. In aquatic systems, the concentration of dissolved oxygen is related to water temperature and the speed at which the water moves. Cold water has more dissolved oxygen than warmer water. In addition, salinity, current, and tide can be important abiotic factors in aquatic ecosystems.

## Other Terrestrial Factors

Wind can be an important abiotic factor because it influences the rate of evaporation and transpiration. The physical force of wind is also important because it can move soil, water, or other abiotic factors, as well as an ecosystem's organisms.

Fire is another terrestrial factor that can be an important agent of disturbance in terrestrial ecosystems. Some organisms are adapted to fire and, thus, require the high heat associated with fire to complete a part of their life cycle. For example, the jack pine—a coniferous tree—requires heat from fire for its seed cones to open ([\[link\]](#)). Through the burning of pine needles, fire adds nitrogen to the soil and limits competition by destroying undergrowth.



The mature cones of the jack pine (*Pinus banksiana*) open only when exposed to high temperatures, such as during a forest fire. A fire is likely to kill most vegetation, so a seedling that germinates after a fire is more likely to receive ample sunlight than one that germinates under normal conditions. (credit: USDA)

## Abiotic Factors Influencing Plant Growth

Temperature and moisture are important influences on plant production (primary productivity) and the amount of organic matter available as food (net primary productivity). **Net primary productivity** is an estimation of all of the organic matter available as food; it is calculated as the total amount of carbon fixed per year minus the amount that is oxidized during cellular respiration. In terrestrial environments, net primary productivity is estimated by measuring the **aboveground biomass** per unit area, which is the total mass of living plants, excluding roots. This means that a large percentage of plant biomass which exists underground is not included in this measurement. Net primary productivity is an important variable when considering differences in biomes. Very productive biomes have a high level of aboveground biomass.

Annual biomass production is directly related to the abiotic components of the environment. Environments with the greatest amount of biomass have conditions in which photosynthesis, plant growth, and the resulting net primary productivity are optimized. The climate of these areas is warm and wet. Photosynthesis can proceed at a high rate, enzymes can work most efficiently, and stomata can remain open without the risk of excessive transpiration; together, these factors lead to the maximal amount of carbon dioxide (CO<sub>2</sub>) moving into the plant, resulting in high biomass production. The aboveground biomass produces several important resources for other living things, including habitat and food. Conversely, dry and cold environments have lower photosynthetic rates and therefore less biomass. The animal communities living there will also be affected by the decrease in available food.

## Section Summary

Biogeography is the study of the geographic distribution of living things and the abiotic factors that affect their distribution. Endemic species are species that are naturally found only in a specific geographic area. The



distribution of living things is influenced by several environmental factors that are, in part, controlled by the latitude or elevation at which an organism is found. Ocean upwelling and spring and fall turnovers are important processes regulating the distribution of nutrients and other abiotic factors important in aquatic ecosystems. Energy sources, temperature, water, inorganic nutrients, and soil are factors limiting the distribution of living things in terrestrial systems. Net primary productivity is a measure of the amount of biomass produced by a biome.

## Art Connections

### Exercise:

#### Problem:

[\[link\]](#) How might turnover in tropical lakes differ from turnover in lakes that exist in temperate regions?

---

#### Solution:

[\[link\]](#) Tropical lakes don't freeze, so they don't undergo spring turnover in the same way temperate lakes do. However, stratification does occur, as well as seasonal turnover.

## Review Questions

### Exercise:

#### Problem:

Understory plants in a temperate forest have adaptations to capture limited \_\_\_\_\_.

- a. water
- b. nutrients
- c. heat
- d. sunlight

---

**Solution:**

D

**Exercise:****Problem:**

An ecologist hiking up a mountain may notice different biomes along the way due to changes in all of the following except:

- a. elevation
- b. rainfall
- c. latitude
- d. temperature

---

**Solution:**

C

**Free Response****Exercise:****Problem:**

Compare and contrast ocean upwelling and spring and fall turnovers.

---

**Solution:**

Ocean upwelling is a continual process that occurs year-round. Spring and fall turnover in freshwater lakes and ponds, however, is a seasonal process that occurs due to temperature changes in the water that take place during springtime warming and autumn cooling. Both ocean upwelling and spring and fall turnover enable nutrients in the organic materials at the bottom of the body of water to be recycled and reused by living things.

**Exercise:****Problem:**

Many endemic species are found in areas that are geographically isolated. Suggest a plausible scientific explanation for why this is so.

---

**Solution:**

Areas that have been geographically isolated for very long periods of time allow unique species to evolve; these species are distinctly different from those of surrounding areas and remain so, since geographic isolation keeps them separated from other species.

**Glossary**

aboveground biomass

total mass of aboveground living plants per area

biogeography

study of the geographic distribution of living things and the abiotic factors that affect their distribution

biome

ecological community of plants, animals, and other organisms that is adapted to a characteristic set of environmental conditions

endemic

species found only in a specific geographic area that is usually restricted in size

fall and spring turnover

seasonal process that recycles nutrients and oxygen from the bottom of a freshwater ecosystem to the top

net primary productivity

measurement of the energy accumulation within an ecosystem, calculated as the total amount of carbon fixed per year minus the

amount that is oxidized during cellular respiration

ocean upwelling

rising of deep ocean waters that occurs when prevailing winds blow along surface waters near a coastline

thermocline

layer of water with a temperature that is significantly different from that of the surrounding layers

## Terrestrial Biomes

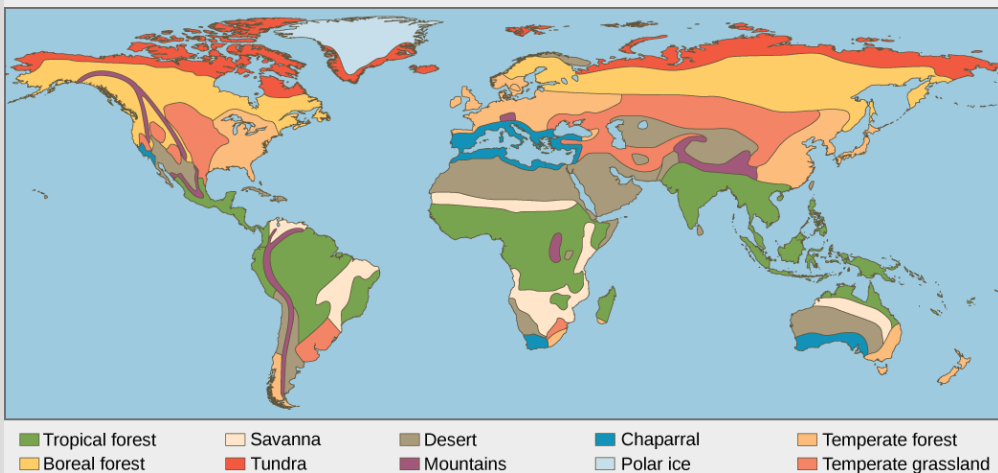
By the end of this section, you will be able to:

- Identify the two major abiotic factors that determine terrestrial biomes
- Recognize distinguishing characteristics of each of the eight major terrestrial biomes

The Earth's biomes are categorized into two major groups: terrestrial and aquatic. Terrestrial biomes are based on land, while aquatic biomes include both ocean and freshwater biomes. The eight major terrestrial biomes on Earth are each distinguished by characteristic temperatures and amount of precipitation. Comparing the annual totals of precipitation and fluctuations in precipitation from one biome to another provides clues as to the importance of abiotic factors in the distribution of biomes. Temperature variation on a daily and seasonal basis is also important for predicting the geographic distribution of the biome and the vegetation type in the biome. The distribution of these biomes shows that the same biome can occur in geographically distinct areas with similar climates ([link](#)).

### Note:

#### Art Connection



Each of the world's major biomes is distinguished by characteristic temperatures and amounts of precipitation. Polar ice and mountains are also shown.

Which of the following statements about biomes is false?

- a. Chaparral is dominated by shrubs.
- b. Savannas and temperate grasslands are dominated by grasses.
- c. Boreal forests are dominated by deciduous trees.
- d. Lichens are common in the arctic tundra.

## **Tropical Wet Forest**

Tropical wet forests are also referred to as tropical rainforests. This biome is found in equatorial regions ([link](#)). The vegetation is characterized by plants with broad leaves that fall off throughout the year. Unlike the trees of deciduous forests, the trees in this biome do not have a seasonal loss of leaves associated with variations in temperature and sunlight; these forests are “evergreen” year-round.

The temperature and sunlight profiles of tropical wet forests are very stable in comparison to that of other terrestrial biomes, with the temperatures ranging from 20 °C to 34 °C (68 °F to 93 °F). When one compares the annual temperature variation of tropical wet forests with that of other forest biomes, the lack of seasonal temperature variation in the tropical wet forest becomes apparent. This lack of seasonality leads to year-round plant growth, rather than the seasonal (spring, summer, and fall) growth seen in other biomes. In contrast to other ecosystems, tropical ecosystems do not have long days and short days during the yearly cycle. Instead, a constant daily amount of sunlight (11–12 hrs per day) provides more solar radiation, thereby, a longer period of time for plant growth.

The annual rainfall in tropical wet forests ranges from 125 to 660 cm (50–200 in) with some monthly variation. While sunlight and temperature remain fairly consistent, annual rainfall is highly variable. Tropical wet forests have wet months in which there can be more than 30 cm (11–12 in) of precipitation, as well as dry months in which there are fewer than 10 cm

(3.5 in) of rainfall. However, the driest month of a tropical wet forest still exceeds the *annual* rainfall of some other biomes, such as deserts.

Tropical wet forests have high net primary productivity because the annual temperatures and precipitation values in these areas are ideal for plant growth. Therefore, the extensive biomass present in the tropical wet forest leads to plant communities with very high species diversities ([\[link\]](#)).

Tropical wet forests have more species of trees than any other biome; on average between 100 and 300 species of trees are present in a single hectare (2.5 acres) of South America. One way to visualize this is to compare the distinctive horizontal layers within the tropical wet forest biome. On the forest floor is a sparse layer of plants and decaying plant matter. Above that is an understory of short shrubby foliage. A layer of trees rises above this understory and is topped by a closed upper **canopy**—the uppermost overhead layer of branches and leaves. Some additional trees emerge through this closed upper canopy. These layers provide diverse and complex habitats for the variety of plants, fungi, animals, and other organisms within the tropical wet forests. For instance, epiphytes are plants that grow on other plants, which typically are not harmed. Epiphytes are found throughout tropical wet forest biomes. Many species of animals use the variety of plants and the complex structure of the tropical wet forests for food and shelter. Some organisms live several meters above ground and have adapted to this arboreal lifestyle.



Tropical wet forests, such as these forests of Madre de Dios, Peru, near the Amazon River, have high species diversity. (credit: Roosevelt Garcia)

## Savannas

Savannas are grasslands with scattered trees, and they are located in Africa, South America, and northern Australia ([\[link\]](#)). Savannas are hot, tropical areas with temperatures averaging from 24 °C to 29 °C (75 °F to 84 °F) and an annual rainfall of 10–40 cm (3.9–15.7 in). Savannas have an extensive dry season; for this reason, forest trees do not grow as well as they do in the tropical wet forest (or other forest biomes). As a result, within the grasses and forbs (herbaceous flowering plants) that dominate the savanna, there are relatively few trees ([\[link\]](#)). Since fire is an important source of disturbance in this biome, plants have evolved well-developed root systems that allow them to quickly re-sprout after a fire.



Savannas, like this one in Taita Hills Wildlife Sanctuary in Kenya, are dominated by grasses. (credit: Christopher T. Cooper)



## Subtropical Deserts

Subtropical deserts exist between 15 ° and 30 ° north and south latitude and are centered on the Tropics of Cancer and Capricorn ([\[link\]](#)). This biome is very dry; in some years, evaporation exceeds precipitation. Subtropical hot deserts can have daytime soil surface temperatures above 60 °C (140 °F) and nighttime temperatures approaching 0 °C (32 °F). In cold deserts, temperatures can be as high as 25 °C and can drop below -30 °C (-22 °F). Subtropical deserts are characterized by low annual precipitation of fewer than 30 cm (12 in) with little monthly variation and lack of predictability in rainfall. In some cases, the annual rainfall can be as low as 2 cm (0.8 in) in subtropical deserts located in central Australia (“the Outback”) and northern Africa.

The vegetation and low animal diversity of this biome is closely related to this low and unpredictable precipitation. Very dry deserts lack perennial vegetation that lives from one year to the next; instead, many plants are annuals that grow quickly and reproduce when rainfall does occur, then they die. Many other plants in these areas are characterized by having a number of adaptations that conserve water, such as deep roots, reduced foliage, and water-storing stems ([\[link\]](#)). Seed plants in the desert produce seeds that can be in dormancy for extended periods between rains. Adaptations in desert animals include nocturnal behavior and burrowing.



To reduce water loss, many desert plants have tiny leaves or no leaves at all. The leaves of ocotillo (*Fouquieria splendens*), shown here in the Sonora Desert near Gila Bend, Arizona, appear only after rainfall, and then are shed.

## Chaparral

The chaparral is also called the scrub forest and is found in California, along the Mediterranean Sea, and along the southern coast of Australia ([link](#)). The annual rainfall in this biome ranges from 65 cm to 75 cm (25.6–29.5 in), and the majority of the rain falls in the winter. Summers are very dry and many chaparral plants are dormant during the summertime.

The chaparral vegetation, shown in [\[link\]](#), is dominated by shrubs and is adapted to periodic fires, with some plants producing seeds that only germinate after a hot fire. The ashes left behind after a fire are rich in nutrients like nitrogen that fertilize the soil and promote plant regrowth.



The chaparral is dominated by shrubs. (credit: Miguel Vieira)

## Temperate Grasslands

Temperate grasslands are found throughout central North America, where they are also known as prairies; they are also in Eurasia, where they are known as steppes ([\[link\]](#)). Temperate grasslands have pronounced annual fluctuations in temperature with hot summers and cold winters. The annual temperature variation produces specific growing seasons for plants. Plant growth is possible when temperatures are warm enough to sustain plant growth and when ample water is available, which occurs in the spring, summer, and fall. During much of the winter, temperatures are low, and water, which is stored in the form of ice, is not available for plant growth.

Annual precipitation ranges from 25 cm to 75 cm (9.8–29.5 in). Because of relatively lower annual precipitation in temperate grasslands, there are few

trees except for those found growing along rivers or streams. The dominant vegetation tends to consist of grasses and some prairies sustain populations of grazing animals [\[link\]](#). The vegetation is very dense and the soils are fertile because the subsurface of the soil is packed with the roots and rhizomes (underground stems) of these grasses. The roots and rhizomes act to anchor plants into the ground and replenish the organic material (humus) in the soil when they die and decay.



The American bison (*Bison bison*), more commonly called the buffalo, is a grazing mammal that once populated American prairies in huge numbers. (credit: Jack Dykinga, USDA Agricultural Research Service)

Fires, mainly caused by lightning, are a natural disturbance in temperate grasslands. When fire is suppressed in temperate grasslands, the vegetation eventually converts to scrub and dense forests. Often, the restoration or management of temperate grasslands requires the use of controlled burns to suppress the growth of trees and maintain the grasses.

## **Temperate Forests**

Temperate forests are the most common biome in eastern North America, Western Europe, Eastern Asia, Chile, and New Zealand ([link](#)). This biome is found throughout mid-latitude regions. Temperatures range between -30 °C and 30 °C (-22 °F to 86 °F) and drop to below freezing on an annual basis. These temperatures mean that temperate forests have defined growing seasons during the spring, summer, and early fall. Precipitation is relatively constant throughout the year and ranges between 75 cm and 150 cm (29.5–59 in).

Because of the moderate annual rainfall and temperatures, deciduous trees are the dominant plant in this biome ([link](#)). Deciduous trees lose their leaves each fall and remain leafless in the winter. Thus, no photosynthesis occurs in the deciduous trees during the dormant winter period. Each spring, new leaves appear as the temperature increases. Because of the dormant period, the net primary productivity of temperate forests is less than that of tropical wet forests. In addition, temperate forests show less diversity of tree species than tropical wet forest biomes.



Deciduous trees are the dominant plant in the temperate forest.  
(credit: Oliver Herold)



The trees of the temperate forests leaf out and shade much of the ground; however, this biome is more open than tropical wet forests because trees in the temperate forests do not grow as tall as the trees in tropical wet forests. The soils of the temperate forests are rich in inorganic and organic nutrients. This is due to the thick layer of leaf litter on forest floors. As this leaf litter decays, nutrients are returned to the soil. The leaf litter also protects soil from erosion, insulates the ground, and provides habitats for invertebrates (such as the pill bug or roly-poly, *Armadillidium vulgare*) and their predators, such as the red-backed salamander (*Plethodon cinereus*).

## **Boreal Forests**

The boreal forest, also known as taiga or coniferous forest, is found south of the Arctic Circle and across most of Canada, Alaska, Russia, and northern Europe ([\[link\]](#)). This biome has cold, dry winters and short, cool, wet summers. The annual precipitation is from 40 cm to 100 cm (15.7–39 in) and usually takes the form of snow. Little evaporation occurs because of the cold temperatures.

The long and cold winters in the boreal forest have led to the predominance of cold-tolerant cone-bearing plants. These are evergreen coniferous trees like pines, spruce, and fir, which retain their needle-shaped leaves year-round. Evergreen trees can photosynthesize earlier in the spring than deciduous trees because less energy from the sun is required to warm a needle-like leaf than a broad leaf. This benefits evergreen trees, which grow faster than deciduous trees in the boreal forest. In addition, soils in boreal forest regions tend to be acidic with little available nitrogen. Leaves are a nitrogen-rich structure and deciduous trees must produce a new set of these nitrogen-rich structures each year. Therefore, coniferous trees that retain nitrogen-rich needles may have a competitive advantage over the broad-leaved deciduous trees.

The net primary productivity of boreal forests is lower than that of temperate forests and tropical wet forests. The aboveground biomass of boreal forests is high because these slow-growing tree species are long lived and accumulate standing biomass over time. Plant species diversity is less than that seen in temperate forests and tropical wet forests. Boreal forests

lack the pronounced elements of the layered forest structure seen in tropical wet forests. The structure of a boreal forest is often only a tree layer and a ground layer ([\[link\]](#)). When conifer needles are dropped, they decompose more slowly than broad leaves; therefore, fewer nutrients are returned to the soil to fuel plant growth.



The boreal forest (taiga) has low lying plants and conifer trees.  
(credit: L.B. Brubaker)

## Arctic Tundra

The Arctic tundra lies north of the subarctic boreal forest and is located throughout the Arctic regions of the northern hemisphere ([\[link\]](#)). The average winter temperature is  $-34^{\circ}\text{C}$  ( $-34^{\circ}\text{F}$ ) and the average summer temperature is from  $3^{\circ}\text{C}$  to  $12^{\circ}\text{C}$  ( $37^{\circ}\text{F}$ – $52^{\circ}\text{F}$ ). Plants in the arctic tundra have a very short growing season of approximately 10–12 weeks. However, during this time, there are almost 24 hours of daylight and plant growth is rapid. The annual precipitation of the Arctic tundra is very low with little annual variation in precipitation. And, as in the boreal forests, there is little evaporation due to the cold temperatures.

Plants in the Arctic tundra are generally low to the ground ([\[link\]](#)). There is little species diversity, low net primary productivity, and low aboveground biomass. The soils of the Arctic tundra may remain in a perennially frozen state referred to as **permafrost**. The permafrost makes it impossible for roots to penetrate deep into the soil and slows the decay of organic matter, which inhibits the release of nutrients from organic matter. During the growing season, the ground of the Arctic tundra can be completely covered with plants or lichens.



Low-growing plants such as shrub willow dominate the tundra landscape, shown here in the Arctic National Wildlife Refuge. (credit: USFWS Arctic National Wildlife Refuge)

**Note:**

Link to Learning





Watch this [Assignment Discovery: Biomes video](#) for an overview of biomes. To explore further, select one of the biomes on the extended playlist: desert, savanna, temperate forest, temperate grassland, tropic, tundra.

## Section Summary

The Earth has terrestrial biomes and aquatic biomes. Aquatic biomes include both freshwater and marine environments. There are eight major terrestrial biomes: tropical wet forests, savannas, subtropical deserts, chaparral, temperate grasslands, temperate forests, boreal forests, and Arctic tundra. The same biome can occur in different geographic locations with similar climates. Temperature and precipitation, and variations in both, are key abiotic factors that shape the composition of animal and plant communities in terrestrial biomes. Some biomes, such as temperate grasslands and temperate forests, have distinct seasons, with cold weather and hot weather alternating throughout the year. In warm, moist biomes, such as the tropical wet forest, net primary productivity is high, as warm temperatures, abundant water, and a year-round growing season fuel plant growth. Other biomes, such as deserts and tundra, have low primary productivity due to extreme temperatures and a shortage of available water.

## Art Connections

### Exercise:

#### Problem:

[\[link\]](#) Which of the following statements about biomes is false?

- a. Chaparral is dominated by shrubs.
  - b. Savannas and temperate grasslands are dominated by grasses.
  - c. Boreal forests are dominated by deciduous trees.
  - d. Lichens are common in the arctic tundra.
- 

**Solution:**

[\[link\]](#) C. Boreal forests are not dominated by deciduous trees.

## Review Questions

**Exercise:**

**Problem:**

Which of the following biomes is characterized by abundant water resources?

- a. deserts
  - b. boreal forests
  - c. savannas
  - d. tropical wet forests
- 

**Solution:**

D

**Exercise:**

**Problem:**

Which of the following biomes is characterized by short growing seasons?

- a. deserts
- b. tropical wet forests
- c. Arctic tundras

d. savannas

---

**Solution:**

C

## Free Response

**Exercise:**

**Problem:**

The extremely low precipitation of subtropical desert biomes might lead one to expect fire to be a major disturbance factor; however, fire is more common in the temperate grassland biome than in the subtropic desert biome. Why is this?

---

**Solution:**

Fire is less common in desert biomes than in temperate grasslands because deserts have low net primary productivity and, thus, very little plant biomass to fuel a fire.

**Exercise:**

**Problem:**

In what ways are the subtropical desert and the arctic tundra similar?

---

**Solution:**

Both the subtropical desert and the arctic tundra have a low supply of water. In the desert, this is due to extremely low precipitation, and in the arctic tundra, much of the water is unavailable to plants because it is frozen. Both the subtropical desert and the arctic tundra have low net primary productivity.

## **Glossary**

canopy

branches and foliage of trees that form a layer of overhead coverage in a forest

permafrost

perennially frozen portion of the Arctic tundra soil

## Aquatic Biomes

By the end of this section, you will be able to:

- Describe the effects of abiotic factors on the composition of plant and animal communities in aquatic biomes
- Compare and contrast the characteristics of the ocean zones
- Summarize the characteristics of standing water and flowing water freshwater biomes

## Abiotic Factors Influencing Aquatic Biomes

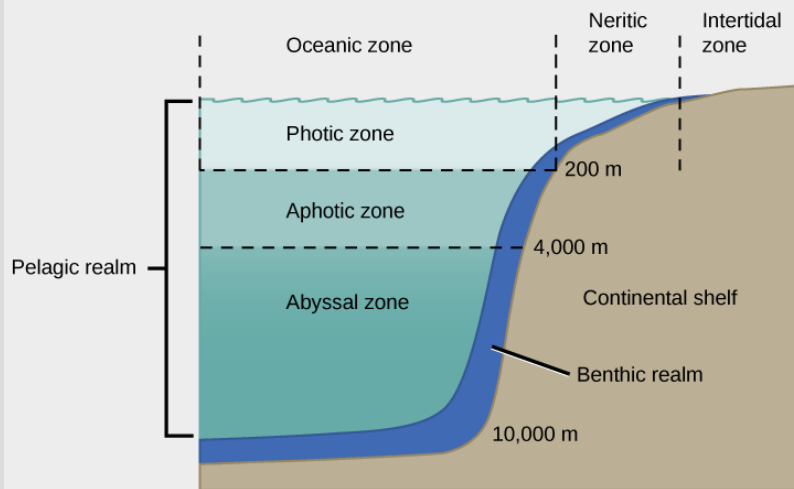
Like terrestrial biomes, aquatic biomes are influenced by a series of abiotic factors. The aquatic medium—water— has different physical and chemical properties than air, however. Even if the water in a pond or other body of water is perfectly clear (there are no suspended particles), water, on its own, absorbs light. As one descends into a deep body of water, there will eventually be a depth which the sunlight cannot reach. While there are some abiotic and biotic factors in a terrestrial ecosystem that might obscure light (like fog, dust, or insect swarms), usually these are not permanent features of the environment. The importance of light in aquatic biomes is central to the communities of organisms found in both freshwater and marine ecosystems. In freshwater systems, stratification due to differences in density is perhaps the most critical abiotic factor and is related to the energy aspects of light. The thermal properties of water (rates of heating and cooling) are significant to the function of marine systems and have major impacts on global climate and weather patterns. Marine systems are also influenced by large-scale physical water movements, such as currents; these are less important in most freshwater lakes.

The ocean is categorized by several areas or zones ([\[link\]](#)). All of the ocean's open water is referred to as the **pelagic realm** (or zone). The **benthic realm** (or zone) extends along the ocean bottom from the shoreline to the deepest parts of the ocean floor. Within the pelagic realm is the **photic zone**, which is the portion of the ocean that light can penetrate (approximately 200 m or 650 ft). At depths greater than 200 m, light cannot penetrate; thus, this is referred to as the **aphotic zone**. The majority of the ocean is aphotic and lacks sufficient light for photosynthesis. The deepest

part of the ocean, the Challenger Deep (in the Mariana Trench, located in the western Pacific Ocean), is about 11,000 m (about 6.8 mi) deep. To give some perspective on the depth of this trench, the ocean is, on average, 4267 m or 14,000 ft deep. These realms and zones are relevant to freshwater lakes as well.

**Note:**

**Art Connection**



The ocean is divided into different zones based on water depth and distance from the shoreline.

In which of the following regions would you expect to find photosynthetic organisms?

- a. the aphotic zone, the neritic zone, the oceanic zone, and the benthic realm
- b. the photic zone, the intertidal zone, the neritic zone, and the oceanic zone
- c. the photic zone, the abyssal zone, the neritic zone, and the oceanic zone
- d. the pelagic realm, the aphotic zone, the neritic zone, and the oceanic zone

## Marine Biomes

The ocean is the largest marine biome. It is a continuous body of salt water that is relatively uniform in chemical composition; it is a weak solution of mineral salts and decayed biological matter. Within the ocean, coral reefs are a second kind of marine biome. Estuaries, coastal areas where salt water and fresh water mix, form a third unique marine biome.

### Ocean

The physical diversity of the ocean is a significant influence on plants, animals, and other organisms. The ocean is categorized into different zones based on how far light reaches into the water. Each zone has a distinct group of species adapted to the biotic and abiotic conditions particular to that zone.

The **intertidal zone**, which is the zone between high and low tide, is the oceanic region that is closest to land ([\[link\]](#)). Generally, most people think of this portion of the ocean as a sandy beach. In some cases, the intertidal zone is indeed a sandy beach, but it can also be rocky or muddy. The intertidal zone is an extremely variable environment because of tides. Organisms are exposed to air and sunlight at low tide and are underwater most of the time, especially during high tide. Therefore, living things that thrive in the intertidal zone are adapted to being dry for long periods of time. The shore of the intertidal zone is also repeatedly struck by waves, and the organisms found there are adapted to withstand damage from the pounding action of the waves ([\[link\]](#)). The exoskeletons of shoreline crustaceans (such as the shore crab, *Carcinus maenas*) are tough and protect them from desiccation (drying out) and wave damage. Another consequence of the pounding waves is that few algae and plants establish themselves in the constantly moving rocks, sand, or mud.



Sea urchins, mussel shells, and starfish are often found in the intertidal zone, shown here in Kachemak Bay, Alaska. (credit: NOAA)

The **neritic zone** ([\[link\]](#)) extends from the intertidal zone to depths of about 200 m (or 650 ft) at the edge of the continental shelf. Since light can penetrate this depth, photosynthesis can occur in the neritic zone. The water here contains silt and is well-oxygenated, low in pressure, and stable in temperature. Phytoplankton and floating *Sargassum* (a type of free-floating marine seaweed) provide a habitat for some sea life found in the neritic zone. Zooplankton, protists, small fishes, and shrimp are found in the neritic zone and are the base of the food chain for most of the world's fisheries.

Beyond the neritic zone is the open ocean area known as the **oceanic zone** ([\[link\]](#)). Within the oceanic zone there is thermal stratification where warm and cold waters mix because of ocean currents. Abundant plankton serve as the base of the food chain for larger animals such as whales and dolphins. Nutrients are scarce and this is a relatively less productive part of the marine biome. When photosynthetic organisms and the protists and animals that feed on them die, their bodies fall to the bottom of the ocean where they remain; unlike freshwater lakes, the open ocean lacks a process for bringing the organic nutrients back up to the surface. The majority of organisms in the aphotic zone include sea cucumbers (phylum



Echinodermata) and other organisms that survive on the nutrients contained in the dead bodies of organisms in the photic zone.

Beneath the pelagic zone is the benthic realm, the deepwater region beyond the continental shelf ([\[link\]](#)). The bottom of the benthic realm is comprised of sand, silt, and dead organisms. Temperature decreases, remaining above freezing, as water depth increases. This is a nutrient-rich portion of the ocean because of the dead organisms that fall from the upper layers of the ocean. Because of this high level of nutrients, a diversity of fungi, sponges, sea anemones, marine worms, sea stars, fishes, and bacteria exist.

The deepest part of the ocean is the **abyssal zone**, which is at depths of 4000 m or greater. The abyssal zone ([\[link\]](#)) is very cold and has very high pressure, high oxygen content, and low nutrient content. There are a variety of invertebrates and fishes found in this zone, but the abyssal zone does not have plants because of the lack of light. Hydrothermal vents are found primarily in the abyssal zone; chemosynthetic bacteria utilize the hydrogen sulfide and other minerals emitted from the vents. These chemosynthetic bacteria use the hydrogen sulfide as an energy source and serve as the base of the food chain found in the abyssal zone.

## Coral Reefs

**Coral reefs** are ocean ridges formed by marine invertebrates living in warm shallow waters within the photic zone of the ocean. They are found within 30° north and south of the equator. The Great Barrier Reef is a well-known reef system located several miles off the northeastern coast of Australia. Other coral reef systems are fringing islands, which are directly adjacent to land, or atolls, which are circular reef systems surrounding a former landmass that is now underwater. The coral organisms (members of phylum Cnidaria) are colonies of saltwater polyps that secrete a calcium carbonate skeleton. These calcium-rich skeletons slowly accumulate, forming the underwater reef ([\[link\]](#)). Corals found in shallower waters (at a depth of approximately 60 m or about 200 ft) have a mutualistic relationship with photosynthetic unicellular algae. The relationship provides corals with the majority of the nutrition and the energy they require. The waters in which

these corals live are nutritionally poor and, without this mutualism, it would not be possible for large corals to grow. Some corals living in deeper and colder water do not have a mutualistic relationship with algae; these corals attain energy and nutrients using stinging cells on their tentacles to capture prey.

**Note:**

Link to Learning



Watch this [National Oceanic and Atmospheric Administration \(NOAA\) video](#) to see marine ecologist Dr. Peter Etnoyer discusses his research on coral organisms.

It is estimated that more than 4,000 fish species inhabit coral reefs. These fishes can feed on coral, the **cryptofauna** (invertebrates found within the calcium carbonate substrate of the coral reefs), or the seaweed and algae that are associated with the coral. In addition, some fish species inhabit the boundaries of a coral reef; these species include **predators**, herbivores, or **planktivores**. Predators are animal species that hunt and are carnivores or “flesh eaters.” Herbivores eat plant material, and planktivores eat plankton.



Coral reefs are formed by the calcium carbonate skeletons of coral organisms, which are marine invertebrates in the phylum Cnidaria.  
(credit: Terry Hughes)

**Note:**

Evolution Connection

**Global Decline of Coral Reefs**

It takes a long time to build a coral reef. The animals that create coral reefs have evolved over millions of years, continuing to slowly deposit the calcium carbonate that forms their characteristic ocean homes. Bathed in warm tropical waters, the coral animals and their symbiotic algal partners evolved to survive at the upper limit of ocean water temperature.

Together, climate change and human activity pose dual threats to the long-term survival of the world's coral reefs. As global warming due to fossil

fuel emissions raises ocean temperatures, coral reefs are suffering. The excessive warmth causes the reefs to expel their symbiotic, food-producing algae, resulting in a phenomenon known as bleaching. When bleaching occurs, the reefs lose much of their characteristic color as the algae and the coral animals die if loss of the symbiotic zooxanthellae is prolonged. Rising levels of atmospheric carbon dioxide further threaten the corals in other ways; as CO<sub>2</sub> dissolves in ocean waters, it lowers the pH and increases ocean acidity. As acidity increases, it interferes with the calcification that normally occurs as coral animals build their calcium carbonate homes.

When a coral reef begins to die, species diversity plummets as animals lose food and shelter. Coral reefs are also economically important tourist destinations, so the decline of coral reefs poses a serious threat to coastal economies.

Human population growth has damaged corals in other ways, too. As human coastal populations increase, the runoff of sediment and agricultural chemicals has increased, too, causing some of the once-clear tropical waters to become cloudy. At the same time, overfishing of popular fish species has allowed the predator species that eat corals to go unchecked. Although a rise in global temperatures of 1–2°C (a conservative scientific projection) in the coming decades may not seem large, it is very significant to this biome. When change occurs rapidly, species can become extinct before evolution leads to new adaptations. Many scientists believe that global warming, with its rapid (in terms of evolutionary time) and inexorable increases in temperature, is tipping the balance beyond the point at which many of the world's coral reefs can recover.

## **Estuaries: Where the Ocean Meets Fresh Water**

**Estuaries** are biomes that occur where a source of fresh water, such as a river, meets the ocean. Therefore, both fresh water and salt water are found in the same vicinity; mixing results in a diluted (brackish) saltwater. Estuaries form protected areas where many of the young offspring of crustaceans, mollusks, and fish begin their lives. Salinity is a very important factor that influences the organisms and the adaptations of the organisms

found in estuaries. The salinity of estuaries varies and is based on the rate of flow of its freshwater sources. Once or twice a day, high tides bring salt water into the estuary. Low tides occurring at the same frequency reverse the current of salt water.

The short-term and rapid variation in salinity due to the mixing of fresh water and salt water is a difficult physiological challenge for the plants and animals that inhabit estuaries. Many estuarine plant species are halophytes: plants that can tolerate salty conditions. Halophytic plants are adapted to deal with the salinity resulting from saltwater on their roots or from sea spray. In some halophytes, filters in the roots remove the salt from the water that the plant absorbs. Other plants are able to pump oxygen into their roots. Animals, such as mussels and clams (phylum Mollusca), have developed behavioral adaptations that expend a lot of energy to function in this rapidly changing environment. When these animals are exposed to low salinity, they stop feeding, close their shells, and switch from aerobic respiration (in which they use gills) to anaerobic respiration (a process that does not require oxygen). When high tide returns to the estuary, the salinity and oxygen content of the water increases, and these animals open their shells, begin feeding, and return to aerobic respiration.

## **Freshwater Biomes**

Freshwater biomes include lakes and ponds (standing water) as well as rivers and streams (flowing water). They also include wetlands, which will be discussed later. Humans rely on freshwater biomes to provide aquatic resources for drinking water, crop irrigation, sanitation, and industry. These various roles and human benefits are referred to as **ecosystem services**. Lakes and ponds are found in terrestrial landscapes and are, therefore, connected with abiotic and biotic factors influencing these terrestrial biomes.

### **Lakes and Ponds**

Lakes and ponds can range in area from a few square meters to thousands of square kilometers. Temperature is an important abiotic factor affecting living things found in lakes and ponds. In the summer, thermal stratification of lakes and ponds occurs when the upper layer of water is warmed by the sun and does not mix with deeper, cooler water. Light can penetrate within the photic zone of the lake or pond. Phytoplankton (algae and cyanobacteria) are found here and carry out photosynthesis, providing the base of the food web of lakes and ponds. Zooplankton, such as rotifers and small crustaceans, consume these phytoplankton. At the bottom of lakes and ponds, bacteria in the aphotic zone break down dead organisms that sink to the bottom.

Nitrogen and phosphorus are important limiting nutrients in lakes and ponds. Because of this, they are determining factors in the amount of phytoplankton growth in lakes and ponds. When there is a large input of nitrogen and phosphorus (from sewage and runoff from fertilized lawns and farms, for example), the growth of algae skyrockets, resulting in a large accumulation of algae called an **algal bloom**. Algal blooms ([\[link\]](#)) can become so extensive that they reduce light penetration in water. As a result, the lake or pond becomes aphotic and photosynthetic plants cannot survive. When the algae die and decompose, severe oxygen depletion of the water occurs. Fishes and other organisms that require oxygen are then more likely to die, and resulting dead zones are found across the globe. Lake Erie and the Gulf of Mexico represent freshwater and marine habitats where phosphorus control and storm water runoff pose significant environmental challenges.



The uncontrolled growth of algae in this lake has resulted in an algal bloom. (credit: Jeremy Nettleton)

## Rivers and Streams

Rivers and streams are continuously moving bodies of water that carry large amounts of water from the source, or headwater, to a lake or ocean. The largest rivers include the Nile River in Africa, the Amazon River in South America, and the Mississippi River in North America.

Abiotic features of rivers and streams vary along the length of the river or stream. Streams begin at a point of origin referred to as **source water**. The source water is usually cold, low in nutrients, and clear. The **channel** (the width of the river or stream) is narrower than at any other place along the length of the river or stream. Because of this, the current is often faster here than at any other point of the river or stream.

The fast-moving water results in minimal silt accumulation at the bottom of the river or stream; therefore, the water is clear. Photosynthesis here is mostly attributed to algae that are growing on rocks; the swift current inhibits the growth of phytoplankton. An additional input of energy can

come from leaves or other organic material that falls into the river or stream from trees and other plants that border the water. When the leaves decompose, the organic material and nutrients in the leaves are returned to the water. Plants and animals have adapted to this fast-moving water. For instance, leeches (phylum Annelida) have elongated bodies and suckers on both ends. These suckers attach to the substrate, keeping the leech anchored in place. Freshwater trout species (phylum Chordata) are an important predator in these fast-moving rivers and streams.

As the river or stream flows away from the source, the width of the channel gradually widens and the current slows. This slow-moving water, caused by the gradient decrease and the volume increase as tributaries unite, has more sedimentation. Phytoplankton can also be suspended in slow-moving water. Therefore, the water will not be as clear as it is near the source. The water is also warmer. Worms (phylum Annelida) and insects (phylum Arthropoda) can be found burrowing into the mud. The higher order predator vertebrates (phylum Chordata) include waterfowl, frogs, and fishes. These predators must find food in these slow moving, sometimes murky, waters and, unlike the trout in the waters at the source, these vertebrates may not be able to use vision as their primary sense to find food. Instead, they are more likely to use taste or chemical cues to find prey.

## **Wetlands**

Wetlands are environments in which the soil is either permanently or periodically saturated with water. Wetlands are different from lakes because wetlands are shallow bodies of water whereas lakes vary in depth.

**Emergent vegetation** consists of wetland plants that are rooted in the soil but have portions of leaves, stems, and flowers extending above the water's surface. There are several types of wetlands including marshes, swamps, bogs, mudflats, and salt marshes ([\[link\]](#)). The three shared characteristics among these types—what makes them wetlands—are their hydrology, hydrophytic vegetation, and hydric soils.





Located in southern Florida, Everglades National Park is vast array of wetland environments, including sawgrass marshes, cypress swamps, and estuarine mangrove forests. Here, a great egret walks among cypress trees. (credit: NPS)

Freshwater marshes and swamps are characterized by slow and steady water flow. Bogs develop in depressions where water flow is low or nonexistent. Bogs usually occur in areas where there is a clay bottom with poor percolation. Percolation is the movement of water through the pores in the soil or rocks. The water found in a bog is stagnant and oxygen depleted because the oxygen that is used during the decomposition of organic matter is not replaced. As the oxygen in the water is depleted, decomposition slows. This leads to organic acids and other acids building up and lowering the pH of the water. At a lower pH, nitrogen becomes unavailable to plants. This creates a challenge for plants because nitrogen is an important limiting resource. Some types of bog plants (such as sundews, pitcher plants, and Venus flytraps) capture insects and extract the nitrogen from their bodies. Bogs have low net primary productivity because the water found in bogs has low levels of nitrogen and oxygen.

## Section Summary

Aquatic ecosystems include both saltwater and freshwater biomes. The abiotic factors important for the structuring of aquatic ecosystems can be different than those seen in terrestrial systems. Sunlight is a driving force behind the structure of forests and also is an important factor in bodies of water, especially those that are very deep, because of the role of photosynthesis in sustaining certain organisms. Density and temperature shape the structure of aquatic systems. Oceans may be thought of as consisting of different zones based on water depth and distance from the shoreline and light penetrance. Different kinds of organisms are adapted to the conditions found in each zone. Coral reefs are unique marine ecosystems that are home to a wide variety of species. Estuaries are found where rivers meet the ocean; their shallow waters provide nourishment and shelter for young crustaceans, mollusks, fishes, and many other species. Freshwater biomes include lakes, ponds, rivers, streams, and wetlands. Bogs are an interesting type of wetland characterized by standing water, lower pH, and a lack of nitrogen.

## Art Connections

### Exercise:

#### Problem:

[\[link\]](#) In which of the following regions would you expect to find photosynthetic organisms?

- a. the aphotic zone, the neritic zone, the oceanic zone, and the benthic realm
- b. the photic zone, the intertidal zone, the neritic zone, and the oceanic zone
- c. the photic zone, the abyssal zone, the neritic zone, and the oceanic zone
- d. the pelagic realm, the aphotic zone, the neritic zone, and the oceanic zone

---

#### Solution:

[\[link\]](#) C. Photosynthetic organisms would be found in the photic, abyssal, neritic, and oceanic zones.

## Review Questions

### Exercise:

#### Problem:

Where would you expect to find the most photosynthesis in an ocean biome?

- a. aphotic zone
- b. abyssal zone
- c. benthic realm
- d. intertidal zone

---

#### Solution:

D

### Exercise:

**Problem:** A key feature of estuaries is:

- a. low light conditions and high productivity
- b. salt water and fresh water
- c. frequent algal blooms
- d. little or no vegetation

---

#### Solution:

B

## Free Response

**Exercise:****Problem:**

Scientists have discovered the bodies of humans and other living things buried in bogs for hundreds of years, but not yet decomposed. Suggest a possible biological explanation for why such bodies are so well-preserved.

---

**Solution:**

Bogs are low in oxygen and high in organic acids. The low oxygen content and the low pH both slow the rate of decomposition.

**Exercise:****Problem:**

Describe the conditions and challenges facing organisms living in the intertidal zone.

---

**Solution:**

Organisms living in the intertidal zone must tolerate periodic exposure to air and sunlight and must be able to be periodically dry. They also must be able to endure the pounding waves; for this reason, some shoreline organisms have hard exoskeletons that provide protection while also reducing the likelihood of drying out.

**Glossary**

abyssal zone

deepest part of the ocean at depths of 4000 m or greater

algal bloom

rapid increase of algae in an aquatic system

aphotic zone

part of the ocean where no light penetrates

benthic realm

(also, benthic zone) part of the ocean that extends along the ocean bottom from the shoreline to the deepest parts of the ocean floor

channel

width of a river or stream from one bank to the other bank

coral reef

ocean ridges formed by marine invertebrates living in warm, shallow waters within the photic zone

cryptofauna

invertebrates found within the calcium carbonate substrate of coral reefs

ecosystem services

human benefits and services provided by natural ecosystems

emergent vegetation

wetland plants that are rooted in the soil but have portions of leaves, stems, and flowers extending above the water's surface

estuary

biomes where a source of fresh water, such as a river, meets the ocean

intertidal zone

part of the ocean that is closest to land; parts extend above the water at low tide

neritic zone

part of the ocean that extends from low tide to the edge of the continental shelf

oceanic zone

part of the ocean that begins offshore where the water measures 200 m deep or deeper

pelagic realm

(also, pelagic zone) open ocean waters that are not close to the bottom or near the shore

photic zone

portion of the ocean that light can penetrate

planktivore

animal species that eats plankton

predator

animal species that hunt and are carnivores or “flesh eaters”

*Sargassum*

type of free-floating marine seaweed

source water

point of origin of a river or stream

## Climate and the Effects of Global Climate Change

By the end of this section, you will be able to:

- Define global climate change
- Summarize the effects of the Industrial Revolution on global atmospheric carbon dioxide concentration
- Describe three natural factors affecting long-term global climate
- List two or more greenhouse gases and describe their role in the greenhouse effect

All biomes are universally affected by global conditions, such as climate, that ultimately shape each biome's environment. Scientists who study climate have noted a series of marked changes that have gradually become increasingly evident during the last sixty years. **Global climate change** is the term used to describe altered global weather patterns, including a worldwide increase in temperature, due largely to rising levels of atmospheric carbon dioxide.

## Climate and Weather

A common misconception about global climate change is that a specific weather event occurring in a particular region (for example, a very cool week in June in central Indiana) is evidence of global climate change. However, a cold week in June is a weather-related event and not a climate-related one. These misconceptions often arise because of confusion over the terms climate and weather.

**Climate** refers to the long-term, predictable atmospheric conditions of a specific area. The climate of a biome is characterized by having consistent temperature and annual rainfall ranges. Climate does not address the amount of rain that fell on one particular day in a biome or the colder-than-average temperatures that occurred on one day. In contrast, **weather** refers to the conditions of the atmosphere during a short period of time. Weather forecasts are usually made for 48-hour cycles. Long-range weather forecasts are available but can be unreliable.

To better understand the difference between climate and weather, imagine that you are planning an outdoor event in northern Wisconsin. You would be thinking about *climate* when you plan the event in the summer rather than the winter because you have long-term knowledge that any given Saturday in the months of May to August would be a better choice for an outdoor event in Wisconsin than any given Saturday in January. However, you cannot determine the specific day that the event should be held on because it is difficult to accurately predict the weather on a specific day. Climate can be considered “average” weather.

## **Global Climate Change**

Climate change can be understood by approaching three areas of study:

- current and past global climate change
- causes of past and present-day global climate change
- ancient and current results of climate change

It is helpful to keep these three different aspects of climate change clearly separated when consuming media reports about global climate change. It is common for reports and discussions about global climate change to confuse the data showing that Earth’s climate is changing with the factors that drive this climate change.

## **Evidence for Global Climate Change**

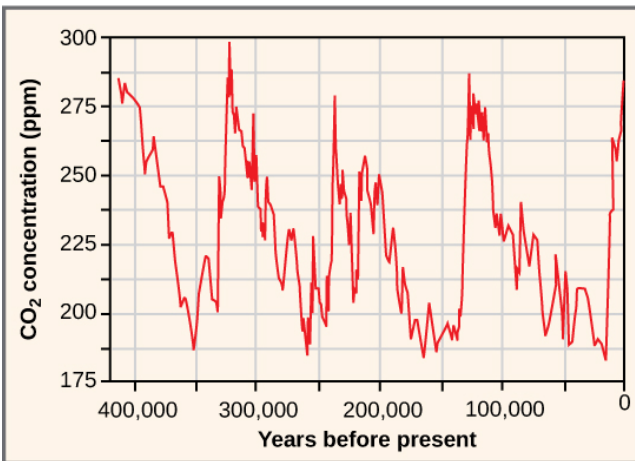
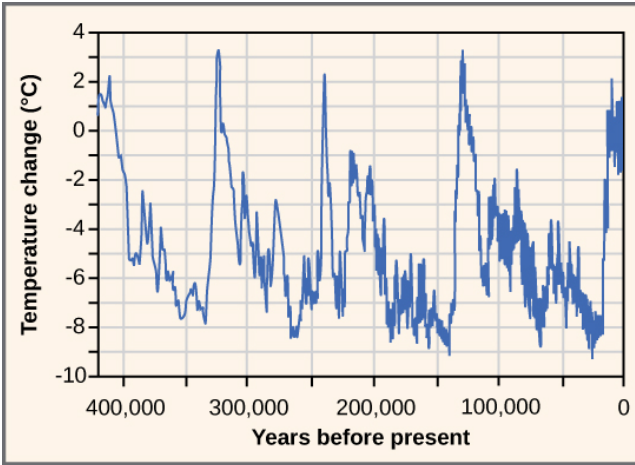
Since scientists cannot go back in time to directly measure climatic variables, such as average temperature and precipitation, they must instead indirectly measure temperature. To do this, scientists rely on historical evidence of Earth’s past climate.

Antarctic ice cores are a key example of such evidence. These ice cores are samples of polar ice obtained by means of drills that reach thousands of meters into ice sheets or high mountain glaciers. Viewing the ice cores is like traveling backwards through time; the deeper the sample, the earlier the time period. Trapped within the ice are bubbles of air and other biological



evidence that can reveal temperature and carbon dioxide data. Antarctic ice cores have been collected and analyzed to indirectly estimate the temperature of the Earth over the past 400,000 years ([\[link\]](#)**a**). The 0 °C on this graph refers to the long-term average. Temperatures that are greater than 0 °C exceed Earth's long-term average temperature. Conversely, temperatures that are less than 0 °C are less than Earth's average temperature. This figure shows that there have been periodic cycles of increasing and decreasing temperature.

Before the late 1800s, the Earth has been as much as 9 °C cooler and about 3 °C warmer. Note that the graph in [\[link\]](#)**b** shows that the atmospheric concentration of carbon dioxide has also risen and fallen in periodic cycles; note the relationship between carbon dioxide concentration and temperature. [\[link\]](#)**b** shows that carbon dioxide levels in the atmosphere have historically cycled between 180 and 300 parts per million (ppm) by volume.

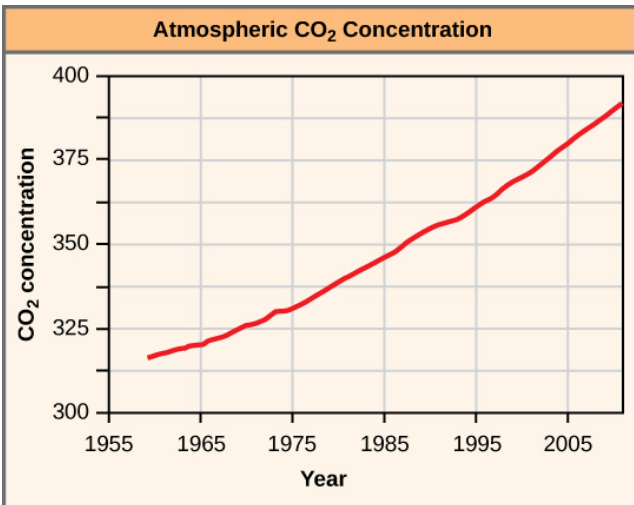


Ice at the Russian Vostok station in East Antarctica was laid down over the course 420,000 years and reached a depth of over 3,000 m. By measuring the amount of CO<sub>2</sub> trapped in the ice, scientists have determined past atmospheric CO<sub>2</sub> concentrations. Temperatures relative to modern day were determined from the amount of deuterium (an isotope of hydrogen) present.

[\[link\]](#) does not show the last 2,000 years with enough detail to compare the changes of Earth's temperature during the last 400,000 years with the temperature change that has occurred in the more recent past. Two significant temperature anomalies, or irregularities, have occurred in the last 2000 years. These are the Medieval Climate Anomaly (or the Medieval Warm Period) and the Little Ice Age. A third temperature anomaly aligns with the Industrial Era. The Medieval Climate Anomaly occurred between 900 and 1300 AD. During this time period, many climate scientists think that slightly warmer weather conditions prevailed in many parts of the world; the higher-than-average temperature changes varied between 0.10 °C and 0.20 °C above the norm. Although 0.10 °C does not seem large enough to produce any noticeable change, it did free seas of ice. Because of this warming, the Vikings were able to colonize Greenland.

The Little Ice Age was a cold period that occurred between 1550 AD and 1850 AD. During this time, a slight cooling of a little less than 1 °C was observed in North America, Europe, and possibly other areas of the Earth. This 1 °C change in global temperature is a seemingly small deviation in temperature (as was observed during the Medieval Climate Anomaly); however, it also resulted in noticeable changes. Historical accounts reveal a time of exceptionally harsh winters with much snow and frost.

The Industrial Revolution, which began around 1750, was characterized by changes in much of human society. Advances in agriculture increased the food supply, which improved the standard of living for people in Europe and the United States. New technologies were invented and provided jobs and cheaper goods. These new technologies were powered using fossil fuels, especially coal. The Industrial Revolution starting in the early nineteenth century ushered in the beginning of the Industrial Era. When a fossil fuel is burned, carbon dioxide is released. With the beginning of the Industrial Era, atmospheric carbon dioxide began to rise ([\[link\]](#)).



The atmospheric concentration of CO<sub>2</sub> has risen steadily since the beginning of industrialization.

## Current and Past Drivers of Global Climate Change

Since it is not possible to go back in time to directly observe and measure climate, scientists use indirect evidence to determine the drivers, or factors, that may be responsible for climate change. The indirect evidence includes data collected using ice cores, boreholes (a narrow shaft bored into the ground), tree rings, glacier lengths, pollen remains, and ocean sediments. The data shows a correlation between the timing of temperature changes and drivers of climate change: before the Industrial Era (pre-1780), there were three drivers of climate change that were not related to human activity or atmospheric gases. The first of these is the Milankovitch cycles. The **Milankovitch cycles** describe the effects of slight changes in the Earth's orbit on Earth's climate. The length of the Milankovitch cycles ranges between 19,000 and 100,000 years. In other words, one could expect to see some predictable changes in the Earth's climate associated with changes in the Earth's orbit at a minimum of every 19,000 years.

The variation in the sun's intensity is the second natural factor responsible for climate change. **Solar intensity** is the amount of solar power or energy the sun emits in a given amount of time. There is a direct relationship between solar intensity and temperature. As solar intensity increases (or decreases), the Earth's temperature correspondingly increases (or decreases). Changes in solar intensity have been proposed as one of several possible explanations for the Little Ice Age.

Finally, volcanic eruptions are a third natural driver of climate change. Volcanic eruptions can last a few days, but the solids and gases released during an eruption can influence the climate over a period of a few years, causing short-term climate changes. The gases and solids released by volcanic eruptions can include carbon dioxide, water vapor, sulfur dioxide, hydrogen sulfide, hydrogen, and carbon monoxide. Generally, volcanic eruptions cool the climate. This occurred in 1783 when volcanos in Iceland erupted and caused the release of large volumes of sulfuric oxide. This led to **haze-effect cooling**, a global phenomenon that occurs when dust, ash, or other suspended particles block out sunlight and trigger lower global temperatures as a result; haze-effect cooling usually extends for one or more years. In Europe and North America, haze-effect cooling produced some of the lowest average winter temperatures on record in 1783 and 1784.

Greenhouse gases are probably the most significant drivers of the climate. When heat energy from the sun strikes the Earth, gases known as **greenhouse gases** trap the heat in the atmosphere, as do the glass panes of a greenhouse keep heat from escaping. The greenhouse gases that affect Earth include carbon dioxide, methane, water vapor, nitrous oxide, and ozone. Approximately half of the radiation from the sun passes through these gases in the atmosphere and strikes the Earth. This radiation is converted into thermal radiation on the Earth's surface, and then a portion of that energy is re-radiated back into the atmosphere. Greenhouse gases, however, reflect much of the thermal energy back to the Earth's surface. The more greenhouse gases there are in the atmosphere, the more thermal energy is reflected back to the Earth's surface. Greenhouse gases absorb and emit radiation and are an important factor in the **greenhouse effect**: the warming

of Earth due to carbon dioxide and other greenhouse gases in the atmosphere.

Evidence supports the relationship between atmospheric concentrations of carbon dioxide and temperature: as carbon dioxide rises, global temperature rises. Since 1950, the concentration of atmospheric carbon dioxide has increased from about 280 ppm to 382 ppm in 2006. In 2011, the atmospheric carbon dioxide concentration was 392 ppm. However, the planet would not be inhabitable by current life forms if water vapor did not produce its drastic greenhouse warming effect.

Scientists look at patterns in data and try to explain differences or deviations from these patterns. The atmospheric carbon dioxide data reveal a historical pattern of carbon dioxide increasing and decreasing, cycling between a low of 180 ppm and a high of 300 ppm. Scientists have concluded that it took around 50,000 years for the atmospheric carbon dioxide level to increase from its low minimum concentration to its higher maximum concentration. However, starting recently, atmospheric carbon dioxide concentrations have increased beyond the historical maximum of 300 ppm. The current increases in atmospheric carbon dioxide have happened very quickly—in a matter of hundreds of years rather than thousands of years. What is the reason for this difference in the rate of change and the amount of increase in carbon dioxide? A key factor that must be recognized when comparing the historical data and the current data is the presence of modern human society; no other driver of climate change has yielded changes in atmospheric carbon dioxide levels at this rate or to this magnitude.

Human activity releases carbon dioxide and methane, two of the most important greenhouse gases, into the atmosphere in several ways. The primary mechanism that releases carbon dioxide is the burning of fossil fuels, such as gasoline, coal, and natural gas ([\[link\]](#)). Deforestation, cement manufacture, animal agriculture, the clearing of land, and the burning of forests are other human activities that release carbon dioxide. Methane ( $\text{CH}_4$ ) is produced when bacteria break down organic matter under anaerobic conditions. Anaerobic conditions can happen when organic matter is trapped underwater (such as in rice paddies) or in the intestines of

herbivores. Methane can also be released from natural gas fields and the decomposition that occurs in landfills. Another source of methane is the melting of clathrates. **Clathrates** are frozen chunks of ice and methane found at the bottom of the ocean. When water warms, these chunks of ice melt and methane is released. As the ocean's water temperature increases, the rate at which clathrates melt is increasing, releasing even more methane. This leads to increased levels of methane in the atmosphere, which further accelerates the rate of global warming. This is an example of the positive feedback loop that is leading to the rapid rate of increase of global temperatures.



The burning of fossil fuels in industry and by vehicles releases carbon dioxide and other greenhouse gases into the atmosphere. (credit: "Pöllö"/Wikimedia Commons)

## Documented Results of Climate Change: Past and Present

Scientists have geological evidence of the consequences of long-ago climate change. Modern-day phenomena such as retreating glaciers and melting polar ice cause a continual rise in sea level. Meanwhile, changes in climate can negatively affect organisms.

### Geological Climate Change

Global warming has been associated with at least one planet-wide extinction event during the geological past. The Permian extinction event occurred about 251 million years ago toward the end of the roughly 50-million-year-long geological time span known as the Permian period. This geologic time period was one of the three warmest periods in Earth's geologic history. Scientists estimate that approximately 70 percent of the terrestrial plant and animal species and 84 percent of marine species became extinct, vanishing forever near the end of the Permian period. Organisms that had adapted to wet and warm climatic conditions, such as annual rainfall of 300–400 cm (118–157 in) and 20 °C–30 °C (68 °F–86 °F) in the tropical wet forest, may not have been able to survive the Permian climate change.

#### Note:

Link to Learning





Watch this [NASA video](#) to discover the mixed effects of global warming on plant growth. While scientists found that warmer temperatures in the 1980s and 1990s caused an increase in plant productivity, this advantage has since been counteracted by more frequent droughts.

## Present Climate Change

A number of global events have occurred that may be attributed to climate change during our lifetimes. Glacier National Park in Montana is undergoing the retreat of many of its glaciers, a phenomenon known as glacier recession. In 1850, the area contained approximately 150 glaciers. By 2010, however, the park contained only about 24 glaciers greater than 25 acres in size. One of these glaciers is the Grinnell Glacier ([link](#)) at Mount Gould. Between 1966 and 2005, the size of Grinnell Glacier shrank by 40 percent. Similarly, the mass of the ice sheets in Greenland and the Antarctic is decreasing: Greenland lost 150–250 km<sup>3</sup> of ice per year between 2002 and 2006. In addition, the size and thickness of the Arctic sea ice is decreasing.



The effect of global warming can be seen in the continuing retreat of Grinnell Glacier. The mean annual temperature in the park has increased 1.33 °C since 1900. The loss of a glacier results in the loss of summer meltwaters, sharply reducing

seasonal water supplies and severely affecting local ecosystems. (credit: modification of work by USGS)

This loss of ice is leading to increases in the global sea level. On average, the sea is rising at a rate of 1.8 mm per year. However, between 1993 and 2010 the rate of sea level increase ranged between 2.9 and 3.4 mm per year. A variety of factors affect the volume of water in the ocean, including the temperature of the water (the density of water is related to its temperature) and the amount of water found in rivers, lakes, glaciers, polar ice caps, and sea ice. As glaciers and polar ice caps melt, there is a significant contribution of liquid water that was previously frozen.

In addition to some abiotic conditions changing in response to climate change, many organisms are also being affected by the changes in temperature. Temperature and precipitation play key roles in determining the geographic distribution and phenology of plants and animals. (Phenology is the study of the effects of climatic conditions on the timing of periodic lifecycle events, such as flowering in plants or migration in birds.) Researchers have shown that 385 plant species in Great Britain are flowering 4.5 days sooner than was recorded earlier during the previous 40 years. In addition, insect-pollinated species were more likely to flower earlier than wind-pollinated species. The impact of changes in flowering date would be mitigated if the insect pollinators emerged earlier. This mismatched timing of plants and pollinators could result in injurious ecosystem effects because, for continued survival, insect-pollinated plants must flower when their pollinators are present.

## **Section Summary**

The Earth has gone through periodic cycles of increases and decreases in temperature. During the past 2000 years, the Medieval Climate Anomaly was a warmer period, while the Little Ice Age was unusually cool. Both of these irregularities can be explained by natural causes of changes in climate, and, although the temperature changes were small, they had significant effects. Natural drivers of climate change include Milankovitch

cycles, changes in solar activity, and volcanic eruptions. None of these factors, however, leads to rapid increases in global temperature or sustained increases in carbon dioxide. The burning of fossil fuels is an important source of greenhouse gases, which plays a major role in the greenhouse effect. Long ago, global warming resulted in the Permian extinction: a large-scale extinction event that is documented in the fossil record. Currently, modern-day climate change is associated with the increased melting of glaciers and polar ice sheets, resulting in a gradual increase in sea level. Plants and animals can also be affected by global climate change when the timing of seasonal events, such as flowering or pollination, is affected by global warming.

## Review Questions

### Exercise:

**Problem:** Which of the following is an example of a weather event?

- a. The hurricane season lasts from June 1 through November 30.
- b. The amount of atmospheric CO<sub>2</sub> has steadily increased during the last century.
- c. A windstorm blew down trees in the Boundary Waters Canoe Area in Minnesota on July 4, 1999.
- d. Deserts are generally dry ecosystems having very little rainfall.

---

### Solution:

C

### Exercise:

### Problem:

Which of the following natural forces is responsible for the release of carbon dioxide and other atmospheric gases?

- a. the Milankovitch cycles

- b. volcanoes
- c. solar intensity
- d. burning of fossil fuels

---

**Solution:**

B

**Free Response**

**Exercise:**

**Problem:**

Compare and contrast how natural- and human-induced processes have influenced global climate change.

---

**Solution:**

Natural processes such as the Milankovitch cycles, variation in solar intensity, and volcanic eruptions can cause periodic, intermittent changes in global climate. Human activity, in the form of emissions from the burning of fossil fuels, has caused a progressive rise in the levels of atmospheric carbon dioxide.

**Exercise:**

**Problem:**

Predict possible consequences if carbon emissions from fossil fuels continue to rise.

---

**Solution:**

If carbon emissions continue to rise, the global temperature will continue to rise; thus, ocean waters will cause the rising of sea levels at the coastlines. Continued melting of glaciers and reduced spring and summer meltwaters may cause summertime water shortages. Changes

in seasonal temperatures may alter lifecycles and interrupt breeding patterns in many species of plants and animals.

## **Glossary**

clathrates

frozen chunks of ice and methane found at the bottom of the ocean

climate

long-term, predictable atmospheric conditions present in a specific area

global climate change

altered global weather patterns, including a worldwide increase in temperature, due largely to rising levels of atmospheric carbon dioxide

greenhouse effect

warming of Earth due to carbon dioxide and other greenhouse gases in the atmosphere

greenhouse gases

atmospheric gases such as carbon dioxide and methane that absorb and emit radiation, thus trapping heat in Earth's atmosphere

haze-effect cooling

effect of the gases and solids from a volcanic eruption on global climate

Milankovitch cycles

cyclic changes in the Earth's orbit that may affect climate

solar intensity

amount of solar power energy the sun emits in a given amount of time

weather

conditions of the atmosphere during a short period of time

## Introduction

class="introduction"

Asian carp  
jump out of  
the water in  
response to  
electrofishing  
. The Asian  
carp in the  
inset  
photograph  
were  
harvested  
from the  
Little  
Calumet  
River in  
Illinois in  
May, 2010,  
using  
rotenone, a  
toxin often  
used as an  
insecticide, in  
an effort to  
learn more  
about the  
population of  
the species.  
(credit main  
image:  
modification  
of work by  
USGS; credit  
inset:  
modification

of work by  
Lt. David  
French,  
USCG)



Imagine sailing down a river in a small motorboat on a weekend afternoon; the water is smooth and you are enjoying the warm sunshine and cool breeze when suddenly you are hit in the head by a 20-pound silver carp. This is a risk now on many rivers and canal systems in Illinois and Missouri because of the presence of Asian carp.

This fish—actually a group of species including the silver, black, grass, and big head carp—has been farmed and eaten in China for over 1000 years. It is one of the most important aquaculture food resources worldwide. In the United States, however, Asian carp is considered a dangerous invasive species that disrupts community structure and composition to the point of threatening native species.

## Population Demography

By the end of this section, you will be able to:

- Describe how ecologists measure population size and density
- Describe three different patterns of population distribution
- Use life tables to calculate mortality rates
- Describe the three types of survivorship curves and relate them to specific populations

Populations are dynamic entities. Populations consist all of the species living within a specific area, and populations fluctuate based on a number of factors: seasonal and yearly changes in the environment, natural disasters such as forest fires and volcanic eruptions, and competition for resources between and within species. The statistical study of population dynamics, **demography**, uses a series of mathematical tools to investigate how populations respond to changes in their biotic and abiotic environments. Many of these tools were originally designed to study human populations. For example, **life tables**, which detail the life expectancy of individuals within a population, were initially developed by life insurance companies to set insurance rates. In fact, while the term “demographics” is commonly used when discussing humans, all living populations can be studied using this approach.

## Population Size and Density

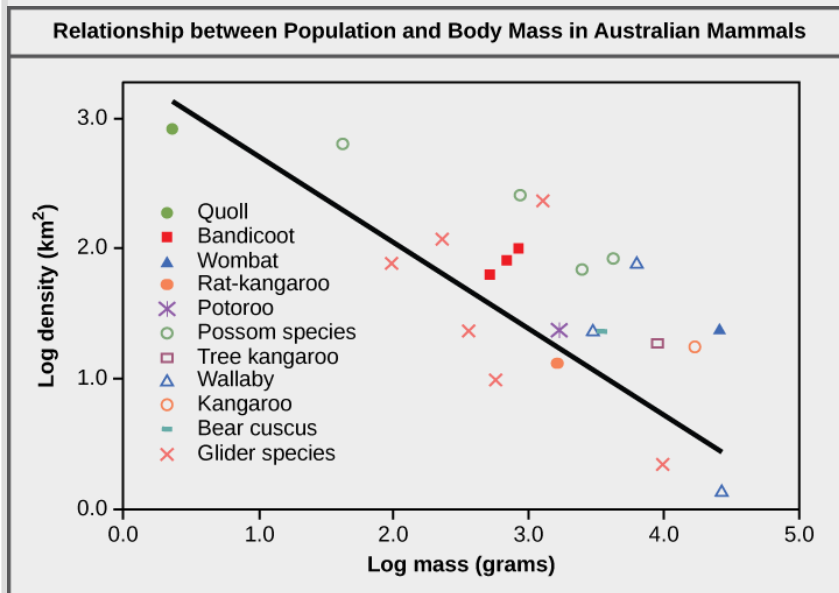
The study of any population usually begins by determining how many individuals of a particular species exist, and how closely associated they are with each other. Within a particular habitat, a population can be characterized by its **population size ( $N$ )**, the total number of individuals, and its **population density**, the number of individuals within a specific area or volume. Population size and density are the two main characteristics used to describe and understand populations. For example, populations with more individuals may be more stable than smaller populations based on their genetic variability, and thus their potential to adapt to the environment. Alternatively, a member of a population with low population density (more spread out in the habitat), might have more difficulty finding a mate to



reproduce compared to a population of higher density. As is shown in [\[link\]](#), smaller organisms tend to be more densely distributed than larger organisms.

### Note:

#### Art Connection



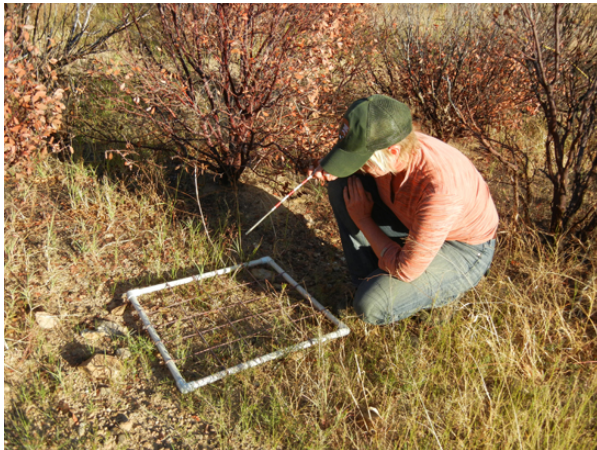
Australian mammals show a typical inverse relationship between population density and body size.

As this graph shows, population density typically decreases with increasing body size. Why do you think this is the case?

### Population Research Methods

The most accurate way to determine population size is to simply count all of the individuals within the habitat. However, this method is often not logistically or economically feasible, especially when studying large habitats. Thus, scientists usually study populations by sampling a representative portion of each habitat and using this data to make inferences

about the habitat as a whole. A variety of methods can be used to sample populations to determine their size and density. For immobile organisms such as plants, or for very small and slow-moving organisms, a **quadrat** may be used ([link](#)). A quadrat is a way of marking off square areas within a habitat, either by staking out an area with sticks and string, or by the use of a wood, plastic, or metal square placed on the ground. After setting the quadrats, researchers then count the number of individuals that lie within their boundaries. Multiple quadrat samples are performed throughout the habitat at several random locations. All of this data can then be used to estimate the population size and population density within the entire habitat. The number and size of quadrat samples depends on the type of organisms under study and other factors, including the density of the organism. For example, if sampling daffodils, a 1 m<sup>2</sup> quadrat might be used whereas with giant redwoods, which are larger and live much further apart from each other, a larger quadrat of 100 m<sup>2</sup> might be employed. This ensures that enough individuals of the species are counted to get an accurate sample that correlates with the habitat, including areas not sampled.



A scientist uses a quadrat to measure population size and density. (credit: NPS Sonoran Desert Network)

For mobile organisms, such as mammals, birds, or fish, a technique called **mark and recapture** is often used. This method involves marking a sample of captured animals in some way (such as tags, bands, paint, or other body markings), and then releasing them back into the environment to allow them to mix with the rest of the population; later, a new sample is collected, including some individuals that are marked (recaptures) and some individuals that are unmarked ([link](#)).



(a)



(b)



(c)

Mark and recapture is used to measure the population size of mobile animals such as (a) bighorn sheep, (b) the California condor, and (c) salmon. (credit a: modification of work by Neal Herbert, NPS; credit b: modification of work by Pacific Southwest Region USFWS; credit c: modification of work by Ingrid Taylar)

Using the ratio of marked and unmarked individuals, scientists determine how many individuals are in the sample. From this, calculations are used to estimate the total population size. This method assumes that the larger the population, the lower the percentage of tagged organisms that will be recaptured since they will have mixed with more untagged individuals. For example, if 80 deer are captured, tagged, and released into the forest, and later 100 deer are captured and 20 of them are already marked, we can determine the population size ( $N$ ) using the following equation:

**Equation:**

$$\frac{(\text{number marked first catch} \times \text{total number of second catch})}{\text{number marked second catch}} = N$$

Using our example, the population size would be estimated at 400.

**Equation:**

$$\frac{(80 \times 100)}{20} = 400$$

Therefore, there are an estimated 400 total individuals in the original population.

There are some limitations to the mark and recapture method. Some animals from the first catch may learn to avoid capture in the second round, thus inflating population estimates. Alternatively, animals may preferentially be retrapped (especially if a food reward is offered), resulting in an underestimate of population size. Also, some species may be harmed by the marking technique, reducing their survival. A variety of other techniques have been developed, including the electronic tracking of animals tagged with radio transmitters and the use of data from commercial fishing and trapping operations to estimate the size and health of populations and communities.

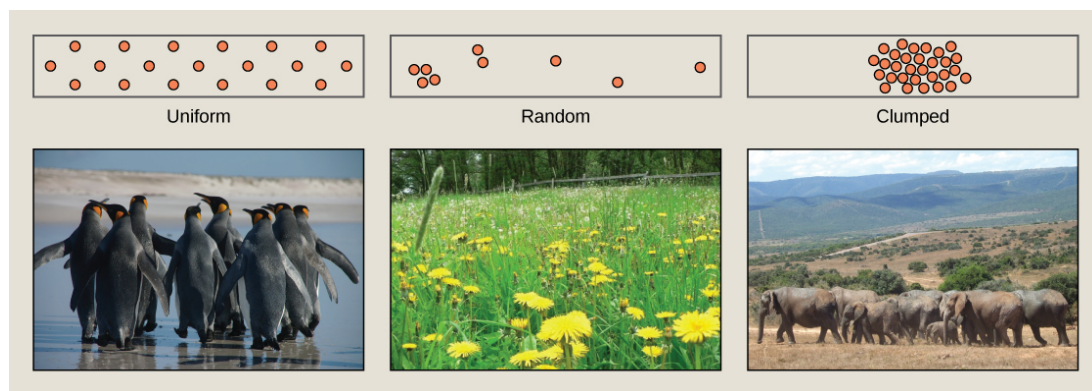
## **Species Distribution**

In addition to measuring simple density, further information about a population can be obtained by looking at the distribution of the individuals.

**Species dispersion patterns** (or distribution patterns) show the spatial relationship between members of a population within a habitat at a particular point in time. In other words, they show whether members of the species live close together or far apart, and what patterns are evident when they are spaced apart.

Individuals in a population can be more or less equally spaced apart, dispersed randomly with no predictable pattern, or clustered in groups. These are known as uniform, random, and clumped dispersion patterns,

respectively ([\[link\]](#)). Uniform dispersion is observed in plants that secrete substances inhibiting the growth of nearby individuals (such as the release of toxic chemicals by the sage plant *Salvia leucophylla*, a phenomenon called allelopathy) and in animals like the penguin that maintain a defined territory. An example of random dispersion occurs with dandelion and other plants that have wind-dispersed seeds that germinate wherever they happen to fall in a favorable environment. A clumped dispersion may be seen in plants that drop their seeds straight to the ground, such as oak trees, or animals that live in groups (schools of fish or herds of elephants). Clumped dispersions may also be a function of habitat heterogeneity. Thus, the dispersion of the individuals within a population provides more information about how they interact with each other than does a simple density measurement. Just as lower density species might have more difficulty finding a mate, solitary species with a random distribution might have a similar difficulty when compared to social species clumped together in groups.



Species may have uniform, random, or clumped distribution.

Territorial birds such as penguins tend to have uniform distribution. Plants such as dandelions with wind-dispersed seeds tend to be randomly distributed. Animals such as elephants that travel in groups exhibit clumped distribution.

(credit a: modification of work by Ben Tubby; credit b: modification of work by Rosendahl; credit c: modification of work by Rebecca Wood)

## Demography

While population size and density describe a population at one particular point in time, scientists must use demography to study the dynamics of a population. Demography is the statistical study of population changes over time: birth rates, death rates, and life expectancies. Each of these measures, especially birth rates, may be affected by the population characteristics described above. For example, a large population size results in a higher birth rate because more potentially reproductive individuals are present. In contrast, a large population size can also result in a higher death rate because of competition, disease, and the accumulation of waste. Similarly, a higher population density or a clumped dispersion pattern results in more potential reproductive encounters between individuals, which can increase birth rate. Lastly, a female-biased sex ratio (the ratio of males to females) or age structure (the proportion of population members at specific age ranges) composed of many individuals of reproductive age can increase birth rates.

In addition, the demographic characteristics of a population can influence how the population grows or declines over time. If birth and death rates are equal, the population remains stable. However, the population size will increase if birth rates exceed death rates; the population will decrease if birth rates are less than death rates. Life expectancy is another important factor; the length of time individuals remain in the population impacts local resources, reproduction, and the overall health of the population. These demographic characteristics are often displayed in the form of a life table.

## Life Tables

Life tables provide important information about the life history of an organism. Life tables divide the population into age groups and often sexes, and show how long a member of that group is likely to live. They are modeled after actuarial tables used by the insurance industry for estimating human life expectancy. Life tables may include the probability of individuals dying before their next birthday (i.e., their **mortality rate**), the

percentage of surviving individuals dying at a particular age interval, and their life expectancy at each interval. An example of a life table is shown in [\[link\]](#) from a study of Dall mountain sheep, a species native to northwestern North America. Notice that the population is divided into age intervals (column A). The mortality rate (per 1000), shown in column D, is based on the number of individuals dying during the age interval (column B) divided by the number of individuals surviving at the beginning of the interval (Column C), multiplied by 1000.

**Equation:**

$$\text{mortality rate} = \frac{\text{number of individuals dying}}{\text{number of individuals surviving}} \times 1000$$

For example, between ages three and four, 12 individuals die out of the 776 that were remaining from the original 1000 sheep. This number is then multiplied by 1000 to get the mortality rate per thousand.

**Equation:**

$$\text{mortality rate} = \frac{12}{776} \times 1000 \approx 15.5$$

As can be seen from the mortality rate data (column D), a high death rate occurred when the sheep were between 6 and 12 months old, and then increased even more from 8 to 12 years old, after which there were few survivors. The data indicate that if a sheep in this population were to survive to age one, it could be expected to live another 7.7 years on average, as shown by the life expectancy numbers in column E.

---

**Life Table of Dall Mountain Sheep**<sup>[\[footnote\]](#)</sup>

**Data Adapted from Edward S. Deevey, Jr., “Life Tables for Natural Populations of Animals,” *The Quarterly Review of Biology* 22, no. 4 (December 1947): 283-314.**

<b>Age interval (years)</b>	<b>Number dying in age interval out of 1000 born</b>	<b>Number surviving at beginning of age interval out of 1000 born</b>	<b>Mortality rate per 1000 alive at beginning of age interval</b>	<b>Life expectancy or mean lifetime remaining to those attaining age interval</b>
0-0.5	54	1000	54.0	7.06
0.5-1	145	946	153.3	--
1-2	12	801	15.0	7.7
2-3	13	789	16.5	6.8
3-4	12	776	15.5	5.9
4-5	30	764	39.3	5.0
5-6	46	734	62.7	4.2
6-7	48	688	69.8	3.4
7-8	69	640	107.8	2.6
8-9	132	571	231.2	1.9
9-10	187	439	426.0	1.3



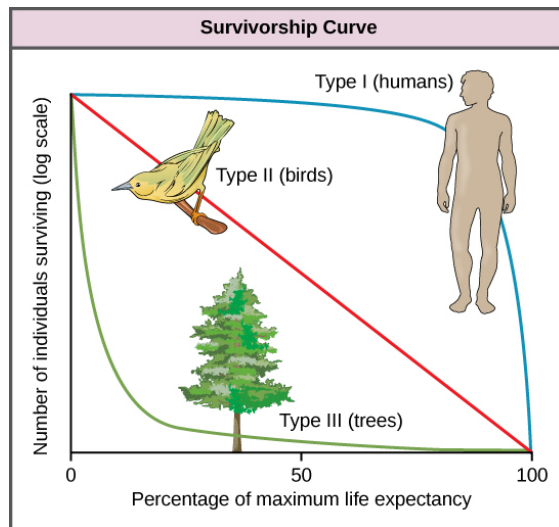
10-11	156	252	619.0	0.9
11-12	90	96	937.5	0.6
12-13	3	6	500.0	1.2
13-14	3	3	1000	0.7

This life table of *Ovis dalli* shows the number of deaths, number of survivors, mortality rate, and life expectancy at each age interval for the Dall mountain sheep.

## Survivorship Curves

Another tool used by population ecologists is a **survivorship curve**, which is a graph of the number of individuals surviving at each age interval plotted versus time (usually with data compiled from a life table). These curves allow us to compare the life histories of different populations ([\[link\]](#)).

Humans and most primates exhibit a Type I survivorship curve because a high percentage of offspring survive their early and middle years—death occurs predominantly in older individuals. These types of species usually have small numbers of offspring at one time, and they give a high amount of parental care to them to ensure their survival. Birds are an example of an intermediate or Type II survivorship curve because birds die more or less equally at each age interval. These organisms also may have relatively few offspring and provide significant parental care. Trees, marine invertebrates, and most fishes exhibit a Type III survivorship curve because very few of these organisms survive their younger years; however, those that make it to an old age are more likely to survive for a relatively long period of time. Organisms in this category usually have a very large number of offspring, but once they are born, little parental care is provided. Thus these offspring are “on their own” and vulnerable to predation, but their sheer numbers assure the survival of enough individuals to perpetuate the species.



Survivorship curves show the distribution of individuals in a population according to age. Humans and most mammals have a Type I survivorship curve because death primarily occurs in the older years. Birds have a Type II survivorship curve, as death at any age is equally probable. Trees have a Type III survivorship curve because very few survive the younger years, but after a certain age, individuals are much more likely to survive.

## Section Summary

Populations are individuals of a species that live in a particular habitat. Ecologists measure characteristics of populations: size, density, dispersion pattern, age structure, and sex ratio. Life tables are useful to calculate life

expectancies of individual population members. Survivorship curves show the number of individuals surviving at each age interval plotted versus time.

## Art Connections

### Exercise:

#### Problem:

[\[link\]](#) As this graph shows, population density typically decreases with increasing body size. Why do you think this is the case?

---

#### Solution:

[\[link\]](#) Smaller animals require less food and other resources, so the environment can support more of them.

## Review Questions

### Exercise:

#### Problem:

Which of the following methods will tell an ecologist about both the size and density of a population?

- a. mark and recapture
  - b. mark and release
  - c. quadrat
  - d. life table
- 

#### Solution:

C

### Exercise:

**Problem:**

Which of the following is best at showing the life expectancy of an individual within a population?

- a. quadrat
- b. mark and recapture
- c. survivorship curve
- d. life table

---

**Solution:**

D

**Exercise:**

**Problem:** Humans have which type of survivorship curve?

- a. Type I
- b. Type II
- c. Type III
- d. Type IV

---

**Solution:**

A

**Free Response****Exercise:****Problem:**

Describe how a researcher would determine the size of a penguin population in Antarctica using the mark and release method.

---

## **Solution:**

The researcher would mark a certain number of penguins with a tag, release them back into the population, and, at a later time, recapture penguins to see what percentage of the recaptured penguins was tagged. This percentage would allow an estimation of the size of the penguin population.

## **Glossary**

demography

statistical study of changes in populations over time

life table

table showing the life expectancy of a population member based on its age

mark and recapture

technique used to determine population size in mobile organisms

mortality rate

proportion of population surviving to the beginning of an age interval that die during the age interval

population density

number of population members divided by the area or volume being measured

population size ( $N$ )

number of population members in a habitat at the same time

quadrat

square made of various materials used to determine population size and density in slow moving or stationary organisms

species dispersion pattern

(also, species distribution pattern) spatial location of individuals of a given species within a habitat at a particular point in time

survivorship curve

graph of the number of surviving population members versus the relative age of the member

## Life Histories and Natural Selection

By the end of this section, you will be able to:

- Describe how life history patterns are influenced by natural selection
- Explain different life history patterns and how different reproductive strategies affect species' survival

A species' **life history** describes the series of events over its lifetime, such as how resources are allocated for growth, maintenance, and reproduction. Life history traits affect the life table of an organism. A species' life history is genetically determined and shaped by the environment and natural selection.

## Life History Patterns and Energy Budgets

Energy is required by all living organisms for their growth, maintenance, and reproduction; at the same time, energy is often a major limiting factor in determining an organism's survival. Plants, for example, acquire energy from the sun via photosynthesis, but must expend this energy to grow, maintain health, and produce energy-rich seeds to produce the next generation. Animals have the additional burden of using some of their energy reserves to acquire food. Furthermore, some animals must expend energy caring for their offspring. Thus, all species have an **energy budget**: they must balance energy intake with their use of energy for metabolism, reproduction, parental care, and energy storage (such as bears building up body fat for winter hibernation).

## Parental Care and Fecundity

**Fecundity** is the potential reproductive capacity of an individual within a population. In other words, fecundity describes how many offspring could ideally be produced if an individual has as many offspring as possible, repeating the reproductive cycle as soon as possible after the birth of the offspring. In animals, fecundity is inversely related to the amount of parental care given to an individual offspring. Species, such as many marine invertebrates, that produce many offspring usually provide little if any care

for the offspring (they would not have the energy or the ability to do so anyway). Most of their energy budget is used to produce many tiny offspring. Animals with this strategy are often self-sufficient at a very early age. This is because of the energy tradeoff these organisms have made to maximize their evolutionary fitness. Because their energy is used for producing offspring instead of parental care, it makes sense that these offspring have some ability to be able to move within their environment and find food and perhaps shelter. Even with these abilities, their small size makes them extremely vulnerable to predation, so the production of many offspring allows enough of them to survive to maintain the species.

Animal species that have few offspring during a reproductive event usually give extensive parental care, devoting much of their energy budget to these activities, sometimes at the expense of their own health. This is the case with many mammals, such as humans, kangaroos, and pandas. The offspring of these species are relatively helpless at birth and need to develop before they achieve self-sufficiency.

Plants with low fecundity produce few energy-rich seeds (such as coconuts and chestnuts) with each having a good chance to germinate into a new organism; plants with high fecundity usually have many small, energy-poor seeds (like orchids) that have a relatively poor chance of surviving. Although it may seem that coconuts and chestnuts have a better chance of surviving, the energy tradeoff of the orchid is also very effective. It is a matter of where the energy is used, for large numbers of seeds or for fewer seeds with more energy.

## **Early versus Late Reproduction**

The timing of reproduction in a life history also affects species survival. Organisms that reproduce at an early age have a greater chance of producing offspring, but this is usually at the expense of their growth and the maintenance of their health. Conversely, organisms that start reproducing later in life often have greater fecundity or are better able to provide parental care, but they risk that they will not survive to reproductive age. Examples of this can be seen in fishes. Small fish like guppies use their



energy to reproduce rapidly, but never attain the size that would give them defense against some predators. Larger fish, like the bluegill or shark, use their energy to attain a large size, but do so with the risk that they will die before they can reproduce or at least reproduce to their maximum. These different energy strategies and tradeoffs are key to understanding the evolution of each species as it maximizes its fitness and fills its niche. In terms of energy budgeting, some species “blow it all” and use up most of their energy reserves to reproduce early before they die. Other species delay having reproduction to become stronger, more experienced individuals and to make sure that they are strong enough to provide parental care if necessary.

### **Single versus Multiple Reproductive Events**

Some life history traits, such as fecundity, timing of reproduction, and parental care, can be grouped together into general strategies that are used by multiple species. **Semelparity** occurs when a species reproduces only once during its lifetime and then dies. Such species use most of their resource budget during a single reproductive event, sacrificing their health to the point that they do not survive. Examples of semelparity are bamboo, which flowers once and then dies, and the Chinook salmon ([link](#)**a**), which uses most of its energy reserves to migrate from the ocean to its freshwater nesting area, where it reproduces and then dies. Scientists have posited alternate explanations for the evolutionary advantage of the Chinook’s post-reproduction death: a programmed suicide caused by a massive release of corticosteroid hormones, presumably so the parents can become food for the offspring, or simple exhaustion caused by the energy demands of reproduction; these are still being debated.

**Iteroparity** describes species that reproduce repeatedly during their lives. Some animals are able to mate only once per year, but survive multiple mating seasons. The pronghorn antelope is an example of an animal that goes into a seasonal estrus cycle (“heat”): a hormonally induced physiological condition preparing the body for successful mating ([link](#)**b**). Females of these species mate only during the estrus phase of the cycle. A different pattern is observed in primates, including humans and

chimpanzees, which may attempt reproduction at any time during their reproductive years, even though their menstrual cycles make pregnancy likely only a few days per month during ovulation ([link](#)c).



The (a) Chinook salmon mates once and dies. The (b) pronghorn antelope mates during specific times of the year during its reproductive life. Primates, such as humans and (c) chimpanzees, may mate on any day, independent of ovulation. (credit a: modification of work by Roger Tabor, USFWS; credit b: modification of work by Mark Gocke, USDA; credit c: modification of work by “Shiny Things”/Flickr)

### Note:

Link to Learning



Play this [interactive PBS evolution-based mating game](#) to learn more about reproductive strategies.

**Note:****Evolution Connection****Energy Budgets, Reproductive Costs, and Sexual Selection in *Drosophila***

Research into how animals allocate their energy resources for growth, maintenance, and reproduction has used a variety of experimental animal models. Some of this work has been done using the common fruit fly, *Drosophila melanogaster*. Studies have shown that not only does reproduction have a cost as far as how long male fruit flies live, but also fruit flies that have already mated several times have limited sperm remaining for reproduction. Fruit flies maximize their last chances at reproduction by selecting optimal mates.

In a 1981 study, male fruit flies were placed in enclosures with either virgin or inseminated females. The males that mated with virgin females had shorter life spans than those in contact with the same number of inseminated females with which they were unable to mate. This effect occurred regardless of how large (indicative of their age) the males were. Thus, males that did not mate lived longer, allowing them more opportunities to find mates in the future.

More recent studies, performed in 2006, show how males select the female with which they will mate and how this is affected by previous matings ([\[link\]](#)).<sup>[footnote]</sup> Males were allowed to select between smaller and larger females. Findings showed that larger females had greater fecundity, producing twice as many offspring per mating as the smaller females did. Males that had previously mated, and thus had lower supplies of sperm, were termed “resource-depleted,” while males that had not mated were termed “non-resource-depleted.” The study showed that although non-resource-depleted males preferentially mated with larger females, this selection of partners was more pronounced in the resource-depleted males. Thus, males with depleted sperm supplies, which were limited in the number of times that they could mate before they replenished their sperm supply, selected larger, more fecund females, thus maximizing their chances for offspring. This study was one of the first to show that the physiological state of the male affected its mating behavior in a way that clearly maximizes its use of limited reproductive resources.

Adapted from Phillip G. Byrne and William R. Rice, “Evidence for adaptive male mate choice in the fruit fly *Drosophila melanogaster*,” Proc

Biol Sci. 273, no. 1589 (2006): 917-922, doi: 10.1098/rspb.2005.3372.

	Ratio large/small females mated
Non sperm-depleted	$8 \pm 5$
Sperm-depleted	$15 \pm 5$

Male fruit flies that had previously mated (sperm-depleted) picked larger, more fecund females more often than those that had not mated (non-sperm-depleted). This change in behavior causes an increase in the efficiency of a limited reproductive resource: sperm.

These studies demonstrate two ways in which the energy budget is a factor in reproduction. First, energy expended on mating may reduce an animal's lifespan, but by this time they have already reproduced, so in the context of natural selection this early death is not of much evolutionary importance. Second, when resources such as sperm (and the energy needed to replenish it) are low, an organism's behavior can change to give them the best chance of passing their genes on to the next generation. These changes in behavior, so important to evolution, are studied in a discipline known as behavioral biology, or ethology, at the interface between population biology and psychology.

## Section Summary

All species have evolved a pattern of living, called a life history strategy, in which they partition energy for growth, maintenance, and reproduction. These patterns evolve through natural selection; they allow species to adapt to their environment to obtain the resources they need to successfully reproduce. There is an inverse relationship between fecundity and parental care. A species may reproduce early in life to ensure surviving to a reproductive age or reproduce later in life to become larger and healthier

and better able to give parental care. A species may reproduce once (semelparity) or many times (iteroparity) in its life.

## Review Questions

### Exercise:

#### Problem:

Which of the following is associated with long-term parental care?

- a. few offspring
- b. many offspring
- c. semelparity
- d. fecundity

---

#### Solution:

A

### Exercise:

#### Problem:

Which of the following is associated with multiple reproductive episodes during a species' lifetime?

- a. semiparity
- b. iteroparity
- c. semelparity
- d. fecundity

---

#### Solution:

B

### Exercise:

**Problem:**

Which of the following is associated with the reproductive potential of a species?

- a. few offspring
- b. many offspring
- c. semelparity
- d. fecundity

---

**Solution:**

D

**Free Response****Exercise:****Problem:**

Why is long-term parental care not associated with having many offspring during a reproductive episode?

---

**Solution:**

Parental care is not feasible for organisms having many offspring because they do not have the energy available to take care of offspring. Most of their energy budget is used in the formation of seeds or offspring, so there is little left for parental care. Also, the sheer number of offspring would make individual parental care impossible.

**Glossary**

energy budget

allocation of energy resources for body maintenance, reproduction, and parental care

fecundity

potential reproductive capacity of an individual

iteroparity

life history strategy characterized by multiple reproductive events during the lifetime of a species

life history

inherited pattern of resource allocation under the influence of natural selection and other evolutionary forces

semelparity

life history strategy characterized by a single reproductive event followed by death

## Environmental Limits to Population Growth

By the end of this section, you will be able to:

- Explain the characteristics of and differences between exponential and logistic growth patterns
- Give examples of exponential and logistic growth in natural populations
- Describe how natural selection and environmental adaptation led to the evolution of particular life history patterns

Although life histories describe the way many characteristics of a population (such as their age structure) change over time in a general way, population ecologists make use of a variety of methods to model population dynamics mathematically. These more precise models can then be used to accurately describe changes occurring in a population and better predict future changes. Certain models that have been accepted for decades are now being modified or even abandoned due to their lack of predictive ability, and scholars strive to create effective new models.

## Exponential Growth

Charles Darwin, in his theory of natural selection, was greatly influenced by the English clergyman Thomas Malthus. Malthus published a book in 1798 stating that populations with unlimited natural resources grow very rapidly, and then population growth decreases as resources become depleted. This accelerating pattern of increasing population size is called **exponential growth**.

The best example of exponential growth is seen in bacteria. Bacteria are prokaryotes that reproduce by prokaryotic fission. This division takes about an hour for many bacterial species. If 1000 bacteria are placed in a large flask with an unlimited supply of nutrients (so the nutrients will not become depleted), after an hour, there is one round of division and each organism divides, resulting in 2000 organisms—an increase of 1000. In another hour, each of the 2000 organisms will double, producing 4000, an increase of 2000 organisms. After the third hour, there should be 8000 bacteria in the flask, an increase of 4000 organisms. The important concept of exponential



growth is that the **population growth rate**—the number of organisms added in each reproductive generation—is accelerating; that is, it is increasing at a greater and greater rate. After 1 day and 24 of these cycles, the population would have increased from 1000 to more than 16 billion. When the population size,  $N$ , is plotted over time, a **J-shaped growth curve** is produced ([link](#)).

The bacteria example is not representative of the real world where resources are limited. Furthermore, some bacteria will die during the experiment and thus not reproduce, lowering the growth rate. Therefore, when calculating the growth rate of a population, the **death rate ( $D$ )** (number organisms that die during a particular time interval) is subtracted from the **birth rate ( $B$ )** (number organisms that are born during that interval). This is shown in the following formula:

**Equation:**

$$\frac{\Delta N \text{ (change in number)}}{\Delta T \text{ (change in time)}} = B \text{ (birth rate)} - D \text{ (death rate)}$$

The birth rate is usually expressed on a per capita (for each individual) basis. Thus,  $B$  (birth rate) =  $bN$  (the per capita birth rate “ $b$ ” multiplied by the number of individuals “ $N$ ”) and  $D$  (death rate) =  $dN$  (the per capita death rate “ $d$ ” multiplied by the number of individuals “ $N$ ”). Additionally, ecologists are interested in the population at a particular point in time, an infinitely small time interval. For this reason, the terminology of differential calculus is used to obtain the “instantaneous” growth rate, replacing the *change* in number and time with an instant-specific measurement of number and time.

**Equation:**

$$\frac{dN}{dT} = bN - dN = (b - d)N$$

Notice that the “ $d$ ” associated with the first term refers to the derivative (as the term is used in calculus) and is different from the death rate, also called “ $d$ .” The difference between birth and death rates is further simplified by

substituting the term “ $r$ ” (intrinsic rate of increase) for the relationship between birth and death rates:

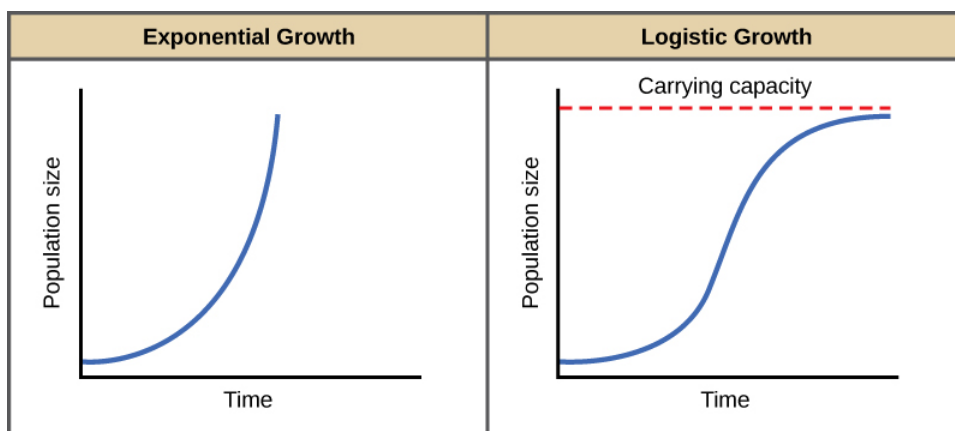
**Equation:**

$$\frac{dN}{dT} = rN$$

The value “ $r$ ” can be positive, meaning the population is increasing in size; or negative, meaning the population is decreasing in size; or zero, where the population’s size is unchanging, a condition known as **zero population growth**. A further refinement of the formula recognizes that different species have inherent differences in their intrinsic rate of increase (often thought of as the potential for reproduction), even under ideal conditions. Obviously, a bacterium can reproduce more rapidly and have a higher intrinsic rate of growth than a human. The maximal growth rate for a species is its **biotic potential, or  $r_{\max}$** , thus changing the equation to:

**Equation:**

$$\frac{dN}{dT} = r_{\max}N$$



When resources are unlimited, populations exhibit exponential growth, resulting in a J-shaped curve.

When resources are limited, populations exhibit

logistic growth. In logistic growth, population expansion decreases as resources become scarce, and it levels off when the carrying capacity of the environment is reached, resulting in an S-shaped curve.

## Logistic Growth

Exponential growth is possible only when infinite natural resources are available; this is not the case in the real world. Charles Darwin recognized this fact in his description of the “struggle for existence,” which states that individuals will compete (with members of their own or other species) for limited resources. The successful ones will survive to pass on their own characteristics and traits (which we know now are transferred by genes) to the next generation at a greater rate (natural selection). To model the reality of limited resources, population ecologists developed the **logistic growth** model.

## Carrying Capacity and the Logistic Model

In the real world, with its limited resources, exponential growth cannot continue indefinitely. Exponential growth may occur in environments where there are few individuals and plentiful resources, but when the number of individuals gets large enough, resources will be depleted, slowing the growth rate. Eventually, the growth rate will plateau or level off ([\[link\]](#)). This population size, which represents the maximum population size that a particular environment can support, is called the **carrying capacity, or  $K$** .

The formula we use to calculate logistic growth adds the carrying capacity as a moderating force in the growth rate. The expression “ $K - N$ ” is indicative of how many individuals may be added to a population at a given stage, and “ $K - N$ ” divided by “ $K$ ” is the fraction of the carrying capacity

available for further growth. Thus, the exponential growth model is restricted by this factor to generate the logistic growth equation:

**Equation:**

$$\frac{dN}{dT} = r_{\max} \frac{dN}{dT} = r_{\max} N \frac{(K - N)}{K}$$

Notice that when  $N$  is very small,  $(K-N)/K$  becomes close to  $K/K$  or 1, and the right side of the equation reduces to  $r_{\max}N$ , which means the population is growing exponentially and is not influenced by carrying capacity. On the other hand, when  $N$  is large,  $(K-N)/K$  come close to zero, which means that population growth will be slowed greatly or even stopped. Thus, population growth is greatly slowed in large populations by the carrying capacity  $K$ . This model also allows for the population of a negative population growth, or a population decline. This occurs when the number of individuals in the population exceeds the carrying capacity (because the value of  $(K-N)/K$  is negative).

A graph of this equation yields an **S-shaped curve** ([\[link\]](#)), and it is a more realistic model of population growth than exponential growth. There are three different sections to an S-shaped curve. Initially, growth is exponential because there are few individuals and ample resources available. Then, as resources begin to become limited, the growth rate decreases. Finally, growth levels off at the carrying capacity of the environment, with little change in population size over time.

## **Role of Intraspecific Competition**

The logistic model assumes that every individual within a population will have equal access to resources and, thus, an equal chance for survival. For plants, the amount of water, sunlight, nutrients, and the space to grow are the important resources, whereas in animals, important resources include food, water, shelter, nesting space, and mates.

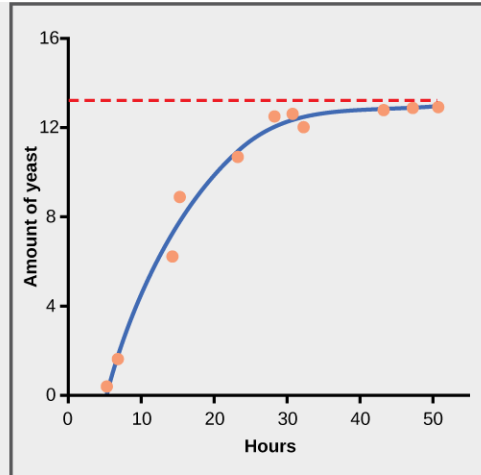
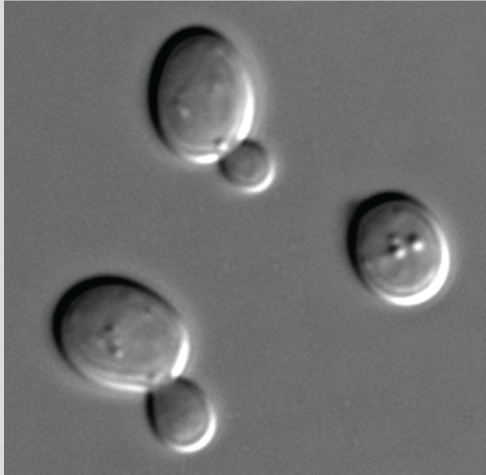
In the real world, phenotypic variation among individuals within a population means that some individuals will be better adapted to their environment than others. The resulting competition between population members of the same species for resources is termed **intraspecific competition** (intra- = “within”; -specific = “species”). Intraspecific competition for resources may not affect populations that are well below their carrying capacity—resources are plentiful and all individuals can obtain what they need. However, as population size increases, this competition intensifies. In addition, the accumulation of waste products can reduce an environment’s carrying capacity.

## Examples of Logistic Growth

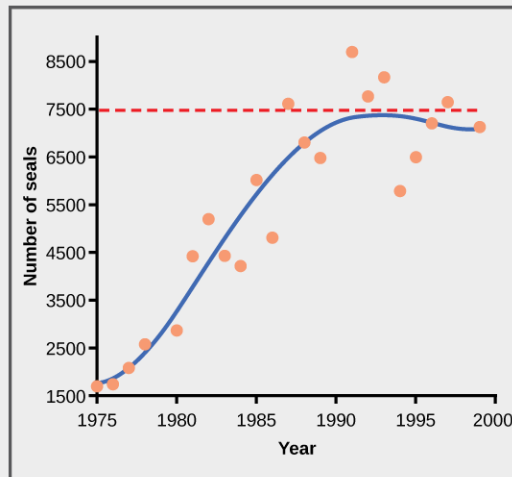
Yeast, a microscopic fungus used to make bread and alcoholic beverages, exhibits the classical S-shaped curve when grown in a test tube ([link](#)a). Its growth levels off as the population depletes the nutrients that are necessary for its growth. In the real world, however, there are variations to this idealized curve. Examples in wild populations include sheep and harbor seals ([link](#)b). In both examples, the population size exceeds the carrying capacity for short periods of time and then falls below the carrying capacity afterwards. This fluctuation in population size continues to occur as the population oscillates around its carrying capacity. Still, even with this oscillation, the logistic model is confirmed.

### Note:

Art Connection



(a)



(b)

(a) Yeast grown in ideal conditions in a test tube show a classical S-shaped logistic growth curve, whereas (b) a natural population of seals shows real-world fluctuation.

If the major food source of the seals declines due to pollution or overfishing, which of the following would likely occur?

- The carrying capacity of seals would decrease, as would the seal population.
- The carrying capacity of seals would decrease, but the seal population would remain the same.

- c. The number of seal deaths would increase but the number of births would also increase, so the population size would remain the same.
- d. The carrying capacity of seals would remain the same, but the population of seals would decrease.

## Section Summary

Populations with unlimited resources grow exponentially, with an accelerating growth rate. When resources become limiting, populations follow a logistic growth curve. The population of a species will level off at the carrying capacity of its environment.

## Art Connections

### Exercise:

#### Problem:

[\[link\]](#) **b** If the major food source of the seals declines due to pollution or overfishing, which of the following would likely occur?

- a. The carrying capacity of seals would decrease, as would the seal population.
- b. The carrying capacity of seals would decrease, but the seal population would remain the same.
- c. The number of seal deaths would increase but the number of births would also increase, so the population size would remain the same.
- d. The carrying capacity of seals would remain the same, but the population of seals would decrease.

---

### Solution:

[\[link\]](#) **b** A

## Review Questions

### Exercise:

#### Problem:

Species with limited resources usually exhibit a(n) \_\_\_\_\_ growth curve.

- a. logistic
- b. logical
- c. experimental
- d. exponential

---

#### Solution:

A

### Exercise:

#### Problem:

The maximum rate of increased characteristic of a species is called its \_\_\_\_\_.

- a. limit
- b. carrying capacity
- c. biotic potential
- d. exponential growth pattern

---

#### Solution:

C

### Exercise:



**Problem:**

The population size of a species capable of being supported by the environment is called its \_\_\_\_\_.

- a. limit
- b. carrying capacity
- c. biotic potential
- d. logistic growth pattern

---

**Solution:**

B

**Free Response****Exercise:****Problem:**

Describe the rate of population growth that would be expected at various parts of the S-shaped curve of logistic growth.

---

**Solution:**

In the first part of the curve, when few individuals of the species are present and resources are plentiful, growth is exponential, similar to a J-shaped curve. Later, growth slows due to the species using up resources. Finally, the population levels off at the carrying capacity of the environment, and it is relatively stable over time.

**Glossary**

biotic potential ( $r_{\max}$ )  
maximal potential growth rate of a species

birth rate ( $B$ )

number of births within a population at a specific point in time

carrying capacity ( $K$ )

number of individuals of a species that can be supported by the limited resources of a habitat

death rate ( $D$ )

number of deaths within a population at a specific point in time

exponential growth

accelerating growth pattern seen in species under conditions where resources are not limiting

intraspecific competition

competition between members of the same species

J-shaped growth curve

shape of an exponential growth curve

logistic growth

leveling off of exponential growth due to limiting resources

population growth rate

number of organisms added in each reproductive generation

S-shaped growth curve

shape of a logistic growth curve

zero population growth

steady population size where birth rates and death rates are equal

## Population Dynamics and Regulation

By the end of this section, you will be able to:

- Give examples of how the carrying capacity of a habitat may change
- Compare and contrast density-dependent growth regulation and density-independent growth regulation, giving examples
- Give examples of exponential and logistic growth in wild animal populations
- Describe how natural selection and environmental adaptation leads to the evolution of particular life-history patterns

The logistic model of population growth, while valid in many natural populations and a useful model, is a simplification of real-world population dynamics. Implicit in the model is that the carrying capacity of the environment does not change, which is not the case. The carrying capacity varies annually: for example, some summers are hot and dry whereas others are cold and wet. In many areas, the carrying capacity during the winter is much lower than it is during the summer. Also, natural events such as earthquakes, volcanoes, and fires can alter an environment and hence its carrying capacity. Additionally, populations do not usually exist in isolation. They engage in **interspecific competition**: that is, they share the environment with other species, competing with them for the same resources. These factors are also important to understanding how a specific population will grow.

Nature regulates population growth in a variety of ways. These are grouped into **density-dependent** factors, in which the density of the population at a given time affects growth rate and mortality, and **density-independent** factors, which influence mortality in a population regardless of population density. Note that in the former, the effect of the factor on the population depends on the density of the population at onset. Conservation biologists want to understand both types because this helps them manage populations and prevent extinction or overpopulation.

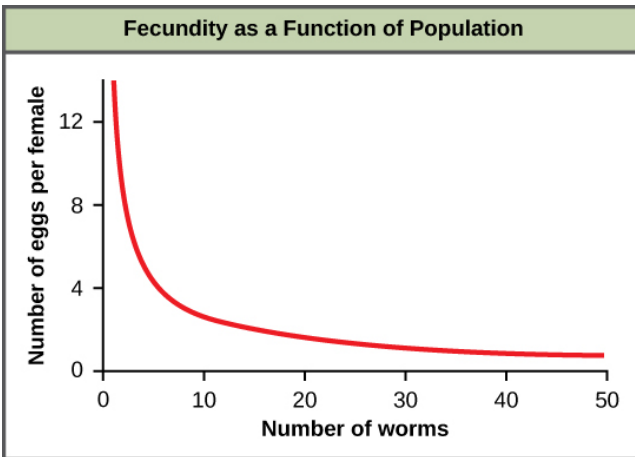
## Density-dependent Regulation

Most density-dependent factors are biological in nature (biotic), and include predation, inter- and intraspecific competition, accumulation of waste, and diseases such as those caused by parasites. Usually, the denser a population is, the greater its mortality rate. For example, during intra- and interspecific competition, the reproductive rates of the individuals will usually be lower, reducing their population's rate of growth. In addition, low prey density increases the mortality of its predator because it has more difficulty locating its food source.

An example of density-dependent regulation is shown in [\[link\]](#) with results from a study focusing on the giant intestinal roundworm (*Ascaris lumbricoides*), a parasite of humans and other mammals. [\[footnote\]](#) Denser populations of the parasite exhibited lower fecundity: they contained fewer eggs. One possible explanation for this is that females would be smaller in more dense populations (due to limited resources) and that smaller females would have fewer eggs. This hypothesis was tested and disproved in a 2009 study which showed that female weight had no influence. [\[footnote\]](#) The actual cause of the density-dependence of fecundity in this organism is still unclear and awaiting further investigation.

N.A. Croll et al., "The Population Biology and Control of *Ascaris lumbricoides* in a Rural Community in Iran." *Transactions of the Royal Society of Tropical Medicine and Hygiene* 76, no. 2 (1982): 187-197, doi:10.1016/0035-9203(82)90272-3.

Martin Walker et al., "Density-Dependent Effects on the Weight of Female *Ascaris lumbricoides* Infections of Humans and its Impact on Patterns of Egg Production." *Parasites & Vectors* 2, no. 11 (February 2009), doi:10.1186/1756-3305-2-11.



In this population of roundworms, fecundity (number of eggs) decreases with population density.  
[\[footnote\]](#)

N.A. Croll et al., "The Population Biology and Control of *Ascaris lumbricoides* in a Rural Community in Iran." *Transactions of the Royal Society of Tropical Medicine and Hygiene* 76, no. 2 (1982): 187-197, doi:10.1016/0035-9203(82)90272-3.

## Density-independent Regulation and Interaction with Density-dependent Factors

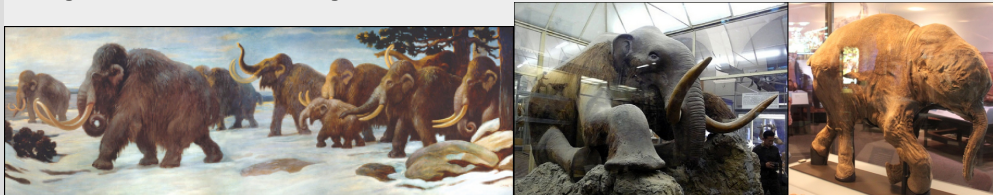
Many factors, typically physical or chemical in nature (abiotic), influence the mortality of a population regardless of its density, including weather, natural disasters, and pollution. An individual deer may be killed in a forest fire regardless of how many deer happen to be in that area. Its chances of survival are the same whether the population density is high or low. The same holds true for cold winter weather.

In real-life situations, population regulation is very complicated and density-dependent and independent factors can interact. A dense population that is reduced in a density-independent manner by some environmental factor(s) will be able to recover differently than a sparse population. For example, a population of deer affected by a harsh winter will recover faster if there are more deer remaining to reproduce.

**Note:**

**Evolution Connection**

**Why Did the Woolly Mammoth Go Extinct?**



The three photos include: (a) 1916 mural of a mammoth herd from the American Museum of Natural History, (b) the only stuffed mammoth in the world, from the Museum of Zoology located in St. Petersburg, Russia, and (c) a one-month-old baby mammoth, named Lyuba, discovered in Siberia in 2007. (credit a: modification of work by Charles R. Knight; credit b: modification of work by “Tanapon”/Flickr; credit c: modification of work by Matt Howry)

It's easy to get lost in the discussion of dinosaurs and theories about why they went extinct 65 million years ago. Was it due to a meteor slamming into Earth near the coast of modern-day Mexico, or was it from some long-term weather cycle that is not yet understood? One hypothesis that will never be proposed is that humans had something to do with it. Mammals were small, insignificant creatures of the forest 65 million years ago, and no humans existed.

Woolly mammoths, however, began to go extinct about 10,000 years ago, when they shared the Earth with humans who were no different

anatomically than humans today ([\[link\]](#)). Mammoths survived in isolated island populations as recently as 1700 BC. We know a lot about these animals from carcasses found frozen in the ice of Siberia and other regions of the north. Scientists have sequenced at least 50 percent of its genome and believe mammoths are between 98 and 99 percent identical to modern elephants.

It is commonly thought that climate change and human hunting led to their extinction. A 2008 study estimated that climate change reduced the mammoth's range from 3,000,000 square miles 42,000 years ago to 310,000 square miles 6,000 years ago. [\[footnote\]](#) It is also well documented that humans hunted these animals. A 2012 study showed that no single factor was exclusively responsible for the extinction of these magnificent creatures. [\[footnote\]](#)

In addition to human hunting, climate change, and reduction of habitat, these scientists demonstrated another important factor in the mammoth's extinction was the migration of humans across the Bering Strait to North America during the last ice age 20,000 years ago.

David Nogués-Bravo et al., "Climate Change, Humans, and the Extinction of the Woolly Mammoth." *PLoS Biol* 6 (April 2008): e79, doi:10.1371/journal.pbio.0060079.

G.M. MacDonald et al., "Pattern of Extinction of the Woolly Mammoth in Beringia." *Nature Communications* 3, no. 893 (June 2012), doi:10.1038/ncomms1881.

The maintenance of stable populations was and is very complex, with many interacting factors determining the outcome. It is important to remember that humans are also part of nature. Once we contributed to a species' decline using primitive hunting technology only.

## Life Histories of *K*-selected and *r*-selected Species

While reproductive strategies play a key role in life histories, they do not account for important factors like limited resources and competition. The regulation of population growth by these factors can be used to introduce a classical concept in population biology, that of *K*-selected versus *r*-selected species.

## Early Theories about Life History: *K*-selected and *r*-selected Species

By the second half of the twentieth century, the concept of *K*- and *r*-selected species was used extensively and successfully to study populations. The concept relates not only reproductive strategies, but also to a species' habitat and behavior, especially in the way that they obtain resources and care for their young. It includes length of life and survivorship factors as well. For this analysis, population biologists have grouped species into the two large categories—*K*-selected and *r*-selected—although they are really two ends of a continuum.

***K*-selected species** are species selected by stable, predictable environments. Populations of *K*-selected species tend to exist close to their carrying capacity (hence the term *K*-selected) where intraspecific competition is high. These species have few, large offspring, a long gestation period, and often give long-term care to their offspring (Table B45\_04\_01). While larger in size when born, the offspring are relatively helpless and immature at birth. By the time they reach adulthood, they must develop skills to compete for natural resources. In plants, scientists think of parental care more broadly: how long fruit takes to develop or how long it remains on the plant are determining factors in the time to the next reproductive event. Examples of *K*-selected species are primates (including humans), elephants, and plants such as oak trees ([link](#)a).

Oak trees grow very slowly and take, on average, 20 years to produce their first seeds, known as acorns. As many as 50,000 acorns can be produced by an individual tree, but the germination rate is low as many of these rot or are eaten by animals such as squirrels. In some years, oaks may produce an exceptionally large number of acorns, and these years may be on a two- or three-year cycle depending on the species of oak (*r*-selection).

As oak trees grow to a large size and for many years before they begin to produce acorns, they devote a large percentage of their energy budget to growth and maintenance. The tree's height and size allow it to dominate other plants in the competition for sunlight, the oak's primary energy resource. Furthermore, when it does reproduce, the oak produces large, energy-rich seeds that use their energy reserve to become quickly established (*K*-selection).



In contrast, ***r*-selected species** have a large number of small offspring (hence their *r* designation ([link](#))). This strategy is often employed in unpredictable or changing environments. Animals that are *r*-selected do not give long-term parental care and the offspring are relatively mature and self-sufficient at birth. Examples of *r*-selected species are marine invertebrates, such as jellyfish, and plants, such as the dandelion ([link](#)**b**). Dandelions have small seeds that are wind dispersed long distances. Many seeds are produced simultaneously to ensure that at least some of them reach a hospitable environment. Seeds that land in inhospitable environments have little chance for survival since their seeds are low in energy content. Note that survival is not necessarily a function of energy stored in the seed itself.

Characteristics of <i>K</i> -selected and <i>r</i> -selected species	
Characteristics of <i>K</i> -selected species	Characteristics of <i>r</i> -selected species
Mature late	Mature early
Greater longevity	Lower longevity
Increased parental care	Decreased parental care
Increased competition	Decreased competition
Fewer offspring	More offspring
Larger offspring	Smaller offspring



(a) K-selected species



(b) r-selected species

(a) Elephants are considered K-selected species as they live long, mature late, and provide long-term parental care to few offspring. Oak trees produce many offspring that do not receive parental care, but are considered K-selected species based on longevity and late maturation. (b) Dandelions and jellyfish are both considered r-selected species as they mature early, have short lifespans, and produce many offspring that receive no parental care.

## Modern Theories of Life History

The *r*- and *K*-selection theory, although accepted for decades and used for much groundbreaking research, has now been reconsidered, and many population biologists have abandoned or modified it. Over the years, several studies attempted to confirm the theory, but these attempts have largely failed. Many species were identified that did not follow the theory's

predictions. Furthermore, the theory ignored the age-specific mortality of the populations which scientists now know is very important. New **demographic-based models** of life history evolution have been developed which incorporate many ecological concepts included in *r*- and *K*-selection theory as well as population age structure and mortality factors.

## Section Summary

Populations are regulated by a variety of density-dependent and density-independent factors. Species are divided into two categories based on a variety of features of their life history patterns: *r*-selected species, which have large numbers of offspring, and *K*-selected species, which have few offspring. The *r*- and *K*-selection theory has fallen out of use; however, many of its key features are still used in newer, demographically-based models of population dynamics.

## Review Questions

### Exercise:

**Problem:** Species that have many offspring at one time are usually:

- a. *r*-selected
- b. *K*-selected
- c. both *r*- and *K*-selected
- d. not selected

---

### Solution:

A

### Exercise:

**Problem:** A forest fire is an example of \_\_\_\_\_ regulation.

- a. density-dependent

- b. density-independent
- c. *r*-selected
- d. *K*-selected

---

**Solution:**

B

**Exercise:**

**Problem:** Primates are examples of:

- a. density-dependent species
- b. density-independent species
- c. *r*-selected species
- d. *K*-selected species

---

**Solution:**

D

## Free Response

**Exercise:**

**Problem:**

Give an example of how density-dependent and density-independent factors might interact.

---

**Solution:**

If a natural disaster such as a fire happened in the winter, when populations are low, it would have a greater effect on the overall population and its recovery than if the same disaster occurred during the summer, when population levels are high.

## Glossary

demographic-based population model

modern model of population dynamics incorporating many features of the  $r$ - and  $K$ -selection theory

density-dependent regulation

regulation of population that is influenced by population density, such as crowding effects; usually involves biotic factors

density-independent regulation

regulation of populations by factors that operate independent of population density, such as forest fires and volcanic eruptions; usually involves abiotic factors

interspecific competition

competition between species for resources in a shared habitat or environment

$K$ -selected species

species suited to stable environments that produce a few, relatively large offspring and provide parental care

$r$ -selected species

species suited to changing environments that produce many offspring and provide little or no parental care

## Human Population Growth

By the end of this section, you will be able to:

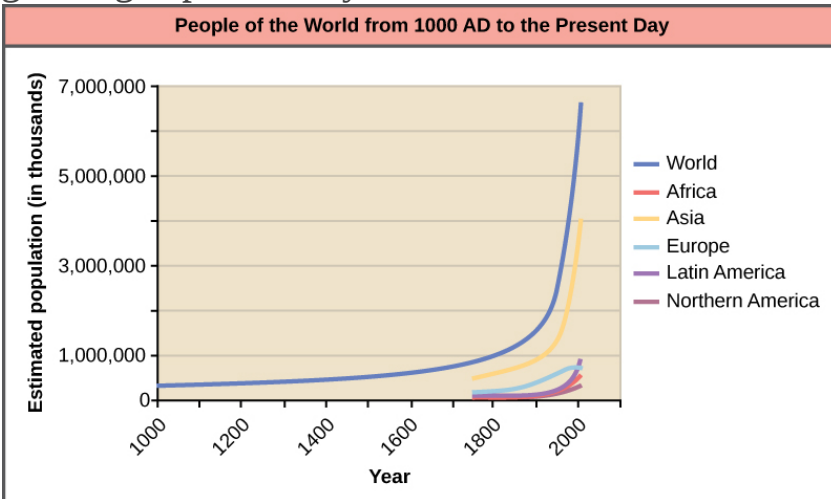
- Discuss how human population growth can be exponential
- Explain how humans have expanded the carrying capacity of their habitat
- Relate population growth and age structure to the level of economic development in different countries
- Discuss the long-term implications of unchecked human population growth

Concepts of animal population dynamics can be applied to human population growth. Humans are not unique in their ability to alter their environment. For example, beaver dams alter the stream environment where they are built. Humans, however, have the ability to alter their environment to increase its carrying capacity sometimes to the detriment of other species (e.g., via artificial selection for crops that have a higher yield). Earth's human population is growing rapidly, to the extent that some worry about the ability of the earth's environment to sustain this population, as long-term exponential growth carries the potential risks of famine, disease, and large-scale death.

Although humans have increased the carrying capacity of their environment, the technologies used to achieve this transformation have caused unprecedented changes to Earth's environment, altering ecosystems to the point where some may be in danger of collapse. The depletion of the ozone layer, erosion due to acid rain, and damage from global climate change are caused by human activities. The ultimate effect of these changes on our carrying capacity is unknown. As some point out, it is likely that the negative effects of increasing carrying capacity will outweigh the positive ones—the carrying capacity of the world for human beings might actually decrease.

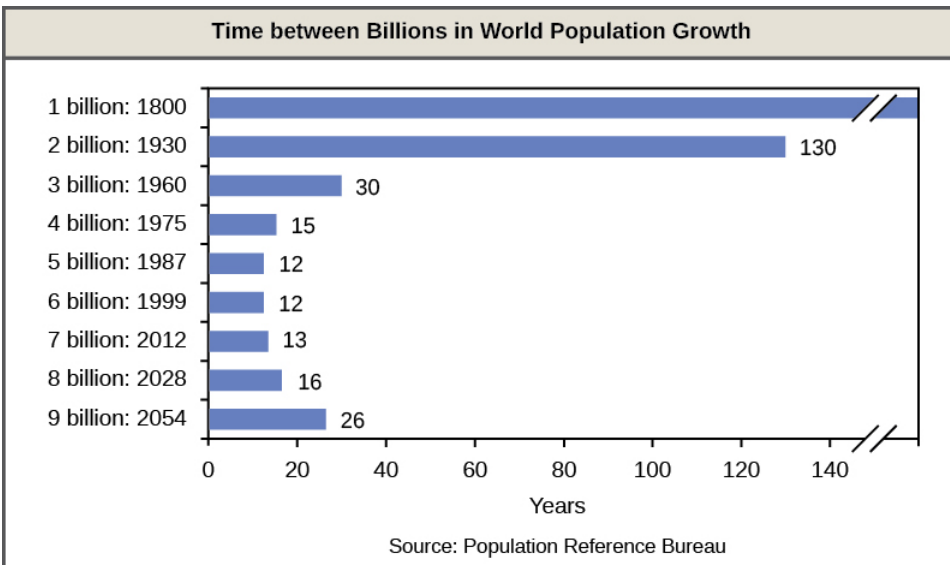
The world's human population is currently experiencing exponential growth even though human reproduction is far below its biotic potential ([\[link\]](#)). To reach its biotic potential, all females would have to become pregnant every nine months or so during their reproductive years. Also, resources would have to be such that the environment would support such growth. Neither of

these two conditions exists. In spite of this fact, human population is still growing exponentially.



Human population growth since 1000 AD is exponential (dark blue line). Notice that while the population in Asia (yellow line), which has many economically underdeveloped countries, is increasing exponentially, the population in Europe (light blue line), where most of the countries are economically developed, is growing much more slowly.

A consequence of exponential human population growth is the time that it takes to add a particular number of humans to the Earth is becoming shorter. [\[link\]](#) shows that 123 years were necessary to add 1 billion humans in 1930, but it only took 24 years to add two billion people between 1975 and 1999. As already discussed, at some point it would appear that our ability to increase our carrying capacity indefinitely on a finite world is uncertain. Without new technological advances, the human growth rate has been predicted to slow in the coming decades. However, the population will still be increasing and the threat of overpopulation remains.



The time between the addition of each billion human beings to Earth decreases over time. (credit: modification of work by Ryan T. Cragun)

**Note:**

Link to Learning



Click through this [interactive view](#) of how human populations have changed over time.

## Overcoming Density-Dependent Regulation



Humans are unique in their ability to alter their environment with the conscious purpose of increasing its carrying capacity. This ability is a major factor responsible for human population growth and a way of overcoming density-dependent growth regulation. Much of this ability is related to human intelligence, society, and communication. Humans can construct shelter to protect them from the elements and have developed agriculture and domesticated animals to increase their food supplies. In addition, humans use language to communicate this technology to new generations, allowing them to improve upon previous accomplishments.

Other factors in human population growth are migration and public health. Humans originated in Africa, but have since migrated to nearly all inhabitable land on the Earth. Public health, sanitation, and the use of antibiotics and vaccines have decreased the ability of infectious disease to limit human population growth. In the past, diseases such as the bubonic plague of the fourteenth century killed between 30 and 60 percent of Europe's population and reduced the overall world population by as many as 100 million people. Today, the threat of infectious disease, while not gone, is certainly less severe. According to the World Health Organization, global death from infectious disease declined from 16.4 million in 1993 to 14.7 million in 1992. To compare to some of the epidemics of the past, the percentage of the world's population killed between 1993 and 2002 decreased from 0.30 percent of the world's population to 0.24 percent. Thus, it appears that the influence of infectious disease on human population growth is becoming less significant.

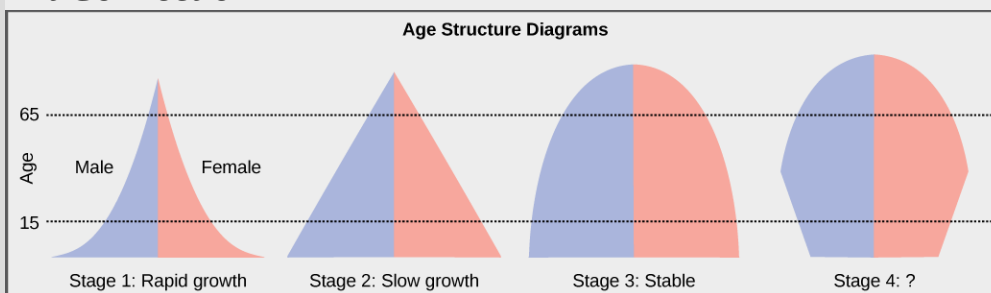
## **Age Structure, Population Growth, and Economic Development**

The age structure of a population is an important factor in population dynamics. **Age structure** is the proportion of a population at different age ranges. Age structure allows better prediction of population growth, plus the ability to associate this growth with the level of economic development in the region. Countries with rapid growth have a pyramidal shape in their age structure diagrams, showing a preponderance of younger individuals, many of whom are of reproductive age or will be soon ([link](#)). This pattern is most often observed in underdeveloped countries where individuals do

not live to old age because of less-than-optimal living conditions. Age structures of areas with slow growth, including developed countries such as the United States, still have a pyramidal structure, but with many fewer young and reproductive-aged individuals and a greater proportion of older individuals. Other developed countries, such as Italy, have zero population growth. The age structure of these populations is more conical, with an even greater percentage of middle-aged and older individuals. The actual growth rates in different countries are shown in [\[link\]](#), with the highest rates tending to be in the less economically developed countries of Africa and Asia.

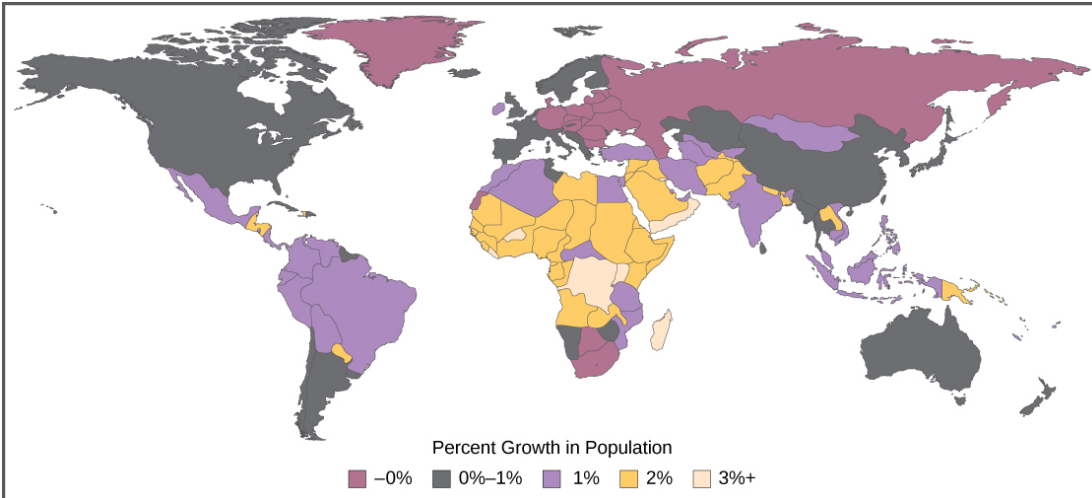
### Note:

#### Art Connection



Typical age structure diagrams are shown. The rapid growth diagram narrows to a point, indicating that the number of individuals decreases rapidly with age. In the slow growth model, the number of individuals decreases steadily with age. Stable population diagrams are rounded on the top, showing that the number of individuals per age group decreases gradually, and then increases for the older part of the population.

Age structure diagrams for rapidly growing, slow growing and stable populations are shown in stages 1 through 3. What type of population change do you think stage 4 represents?



The percent growth rate of population in different countries is shown. Notice that the highest growth is occurring in less economically developed countries in Africa and Asia.

## Long-Term Consequences of Exponential Human Population Growth

Many dire predictions have been made about the world's population leading to a major crisis called the "population explosion." In the 1968 book *The Population Bomb*, biologist Dr. Paul R. Ehrlich wrote, "The battle to feed all of humanity is over. In the 1970s hundreds of millions of people will starve to death in spite of any crash programs embarked upon now. At this late date nothing can prevent a substantial increase in the world death rate."<sup>[footnote]</sup> While many critics view this statement as an exaggeration, the laws of exponential population growth are still in effect, and unchecked human population growth cannot continue indefinitely.

Paul R. Erlich, prologue to *The Population Bomb*, (1968; repr., New York: Ballantine, 1970).

Efforts to control population growth led to the **one-child policy** in China, which used to include more severe consequences, but now imposes fines on urban couples who have more than one child. Due to the fact that some couples wish to have a male heir, many Chinese couples continue to have

more than one child. The policy itself, its social impacts, and the effectiveness of limiting overall population growth are controversial. In spite of population control policies, the human population continues to grow. At some point the food supply may run out because of the subsequent need to produce more and more food to feed our population. The United Nations estimates that future world population growth may vary from 6 billion (a decrease) to 16 billion people by the year 2100. There is no way to know whether human population growth will moderate to the point where the crisis described by Dr. Ehrlich will be averted.

Another result of population growth is the endangerment of the natural environment. Many countries have attempted to reduce the human impact on climate change by reducing their emission of the greenhouse gas carbon dioxide. However, these treaties have not been ratified by every country, and many underdeveloped countries trying to improve their economic condition may be less likely to agree with such provisions if it means slower economic development. Furthermore, the role of human activity in causing climate change has become a hotly debated socio-political issue in some developed countries, including the United States. Thus, we enter the future with considerable uncertainty about our ability to curb human population growth and protect our environment.

**Note:**

Link to Learning



Visit this [website](#) and select “Launch movie” for an animation discussing the global impacts of human population growth.

## Section Summary

The world's human population is growing at an exponential rate. Humans have increased the world's carrying capacity through migration, agriculture, medical advances, and communication. The age structure of a population allows us to predict population growth. Unchecked human population growth could have dire long-term effects on our environment.

## Art Connections

### Exercise:

#### Problem:

[\[link\]](#) Age structure diagrams for rapidly growing, slow growing and stable populations are shown in stages 1 through 3. What type of population change do you think stage 4 represents?

---

#### Solution:

[\[link\]](#) Stage 4 represents a population that is decreasing.

## Review Questions

### Exercise:

#### Problem:

A country with zero population growth is likely to be \_\_\_\_\_.

- a. in Africa
  - b. in Asia
  - c. economically developed
  - d. economically underdeveloped
- 

#### Solution:

D

**Exercise:**

**Problem:**

Which type of country has the greatest proportion of young individuals?

- a. economically developed
- b. economically underdeveloped
- c. countries with zero population growth
- d. countries in Europe

---

**Solution:**

B

**Exercise:**

**Problem:**

Which of the following is not a way that humans have increased the carrying capacity of the environment?

- a. agriculture
- b. using large amounts of natural resources
- c. domestication of animals
- d. use of language

---

**Solution:**

B

**Free Response**

**Exercise:**

**Problem:**

Describe the age structures in rapidly growing countries, slowly growing countries, and countries with zero population growth.

---

**Solution:**

Rapidly growing countries have a large segment of the population at a reproductive age or younger. Slower growing populations have a lower percentage of these individuals, and countries with zero population growth have an even lower percentage. On the other hand, a high proportion of older individuals is seen mostly in countries with zero growth, and a low proportion is most common in rapidly growing countries.

**Glossary**

age structure

proportion of population members at specific age ranges

one-child policy

China's policy to limit population growth by limiting urban couples to have only one child or face the penalty of a fine

## Community Ecology

By the end of this section, you will be able to:

- Discuss the predator-prey cycle
- Give examples of defenses against predation and herbivory
- Describe the competitive exclusion principle
- Give examples of symbiotic relationships between species
- Describe community structure and succession

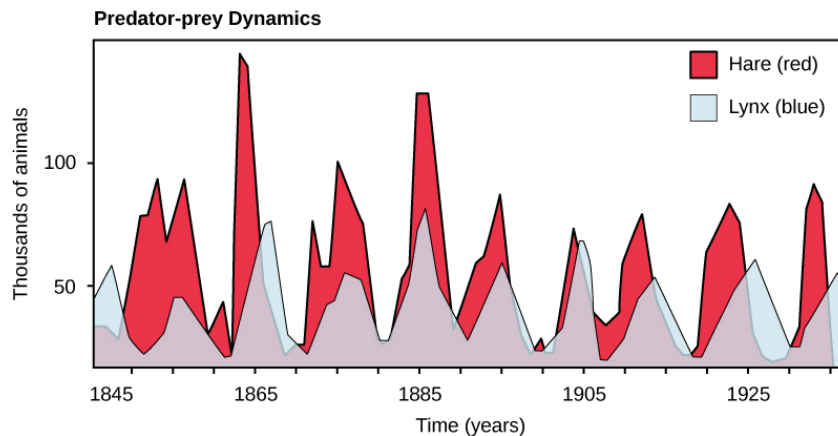
Populations rarely, if ever, live in isolation from populations of other species. In most cases, numerous species share a habitat. The interactions between these populations play a major role in regulating population growth and abundance. All populations occupying the same habitat form a community: populations inhabiting a specific area at the same time. The number of species occupying the same habitat and their relative abundance is known as species diversity. Areas with low diversity, such as the glaciers of Antarctica, still contain a wide variety of living things, whereas the diversity of tropical rainforests is so great that it cannot be counted. Ecology is studied at the community level to understand how species interact with each other and compete for the same resources.

## Predation and Herbivory

Perhaps the classical example of species interaction is predation: the hunting of prey by its predator. Nature shows on television highlight the drama of one living organism killing another. Populations of predators and prey in a community are not constant over time: in most cases, they vary in cycles that appear to be related. The most often cited example of predator-prey dynamics is seen in the cycling of the lynx (predator) and the snowshoe hare (prey), using nearly 200 year-old trapping data from North American forests ([link](#)). This cycle of predator and prey lasts approximately 10 years, with the predator population lagging 1–2 years behind that of the prey population. As the hare numbers increase, there is more food available for the lynx, allowing the lynx population to increase as well. When the lynx population grows to a threshold level, however, they kill so many hares that hare population begins to decline, followed by a decline in the lynx population because of scarcity of food. When the lynx



population is low, the hare population size begins to increase due, at least in part, to low predation pressure, starting the cycle anew.



The cycling of lynx and snowshoe hare populations in Northern Ontario is an example of predator-prey dynamics.

The idea that the population cycling of the two species is entirely controlled by predation models has come under question. More recent studies have pointed to undefined density-dependent factors as being important in the cycling, in addition to predation. One possibility is that the cycling is inherent in the hare population due to density-dependent effects such as lower fecundity (maternal stress) caused by crowding when the hare population gets too dense. The hare cycling would then induce the cycling of the lynx because it is the lynxes' major food source. The more we study communities, the more complexities we find, allowing ecologists to derive more accurate and sophisticated models of population dynamics.

Herbivory describes the consumption of plants by insects and other animals, and it is another interspecific relationship that affects populations. Unlike animals, most plants cannot outrun predators or use mimicry to hide from hungry animals. Some plants have developed mechanisms to defend against herbivory. Other species have developed mutualistic relationships; for

example, herbivory provides a mechanism of seed distribution that aids in plant reproduction.

## **Defense Mechanisms against Predation and Herbivory**

The study of communities must consider evolutionary forces that act on the members of the various populations contained within it. Species are not static, but slowly changing and adapting to their environment by natural selection and other evolutionary forces. Species have evolved numerous mechanisms to escape predation and herbivory. These defenses may be mechanical, chemical, physical, or behavioral.

Mechanical defenses, such as the presence of thorns on plants or the hard shell on turtles, discourage animal predation and herbivory by causing physical pain to the predator or by physically preventing the predator from being able to eat the prey. Chemical defenses are produced by many animals as well as plants, such as the foxglove which is extremely toxic when eaten. [\[link\]](#) shows some organisms' defenses against predation and herbivory.



(a)



(b)



(c)



(d)

The (a) honey locust tree (*Gleditsia triacanthos*) uses thorns, a mechanical defense, against herbivores, while the (b) Florida red-bellied turtle (*Pseudemys nelsoni*) uses its shell as a mechanical defense against predators. (c) Foxglove (*Digitalis* sp.) uses a chemical defense: toxins produced by the plant can cause nausea, vomiting, hallucinations, convulsions, or death when consumed. (d) The North American millipede (*Narceus americanus*) uses both mechanical and chemical defenses: when threatened, the millipede curls into a defensive ball and produces a noxious substance that irritates eyes and skin. (credit a: modification of work by Huw Williams; credit b: modification of work by “JamieS93”/Flickr; credit c: modification of work by Philip Jägenstedt; credit d: modification of work by Cory Zanker)

Many species use their body shape and coloration to avoid being detected by predators. The tropical walking stick is an insect with the coloration and body shape of a twig which makes it very hard to see when stationary against a background of real twigs ([link](#)a). In another example, the chameleon can change its color to match its surroundings ([link](#)b). Both of these are examples of **camouflage**, or avoiding detection by blending in with the background.



(a) The tropical walking stick and (b) the chameleon use body shape and/or coloration to prevent detection by predators. (credit a: modification of work by Linda Tanner; credit b: modification of work by Frank Vassen)

Some species use coloration as a way of warning predators that they are not good to eat. For example, the cinnabar moth caterpillar, the fire-bellied toad, and many species of beetle have bright colors that warn of a foul taste, the presence of toxic chemical, and/or the ability to sting or bite, respectively. Predators that ignore this coloration and eat the organisms will experience their unpleasant taste or presence of toxic chemicals and learn not to eat them in the future. This type of defensive mechanism is called **aposematic coloration**, or warning coloration ([link](#)).



(a)



(b)

(a) The strawberry poison dart frog (*Oophaga pumilio*) uses aposematic coloration to warn predators that it is toxic, while the (b) striped skunk (*Mephitis mephitis*) uses aposematic coloration to warn predators of the unpleasant odor it produces. (credit a: modification of work by Jay Iwasaki; credit b: modification of work by Dan Dzurisin)

While some predators learn to avoid eating certain potential prey because of their coloration, other species have evolved mechanisms to mimic this coloration to avoid being eaten, even though they themselves may not be unpleasant to eat or contain toxic chemicals. In **Batesian mimicry**, a harmless species imitates the warning coloration of a harmful one. Assuming they share the same predators, this coloration then protects the harmless ones, even though they do not have the same level of physical or chemical defenses against predation as the organism they mimic. Many insect species mimic the coloration of wasps or bees, which are stinging, venomous insects, thereby discouraging predation ([link](#)).





Batesian mimicry occurs when a harmless species mimics the coloration of a harmful species, as is seen with the (a) bumblebee and (b) bee-like robber fly. (credit a, b: modification of work by Cory Zanker)

In **Müllerian mimicry**, multiple species share the same warning coloration, but all of them actually have defenses. [\[link\]](#) shows a variety of foul-tasting butterflies with similar coloration. In **Emsleyan/Mertensian mimicry**, a deadly prey mimics a less dangerous one, such as the venomous coral snake mimicking the non-venomous milk snake. This type of mimicry is extremely rare and more difficult to understand than the previous two types. For this type of mimicry to work, it is essential that eating the milk snake has unpleasant but not fatal consequences. Then, these predators learn not to eat snakes with this coloration, protecting the coral snake as well. If the snake were fatal to the predator, there would be no opportunity for the predator to learn not to eat it, and the benefit for the less toxic species would disappear.



Several unpleasant-tasting *Heliconius* butterfly species share a similar color pattern with better-tasting varieties, an example of Müllerian mimicry. (credit: Joron M, Papa R, Beltrán M, Chamberlain N, Mavárez J, et al.)

**Note:**

Link to Learning

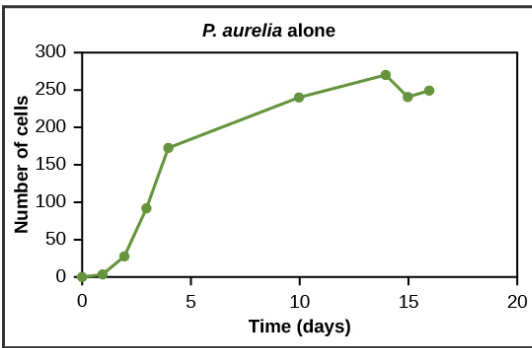


Go to this [website](#) to view stunning examples of mimicry.

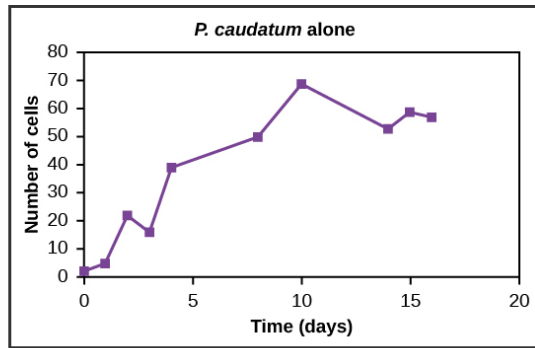
## Competitive Exclusion Principle

Resources are often limited within a habitat and multiple species may compete to obtain them. All species have an ecological niche in the ecosystem, which describes how they acquire the resources they need and how they interact with other species in the community. The **competitive exclusion principle** states that two species cannot occupy the same niche in a habitat. In other words, different species cannot coexist in a community if they are competing for all the same resources. An example of this principle is shown in [\[link\]](#), with two protozoan species, *Paramecium aurelia* and *Paramecium caudatum*. When grown individually in the laboratory, they both thrive. But when they are placed together in the same test tube (habitat), *P. aurelia* outcompetes *P. caudatum* for food, leading to the latter's eventual extinction.

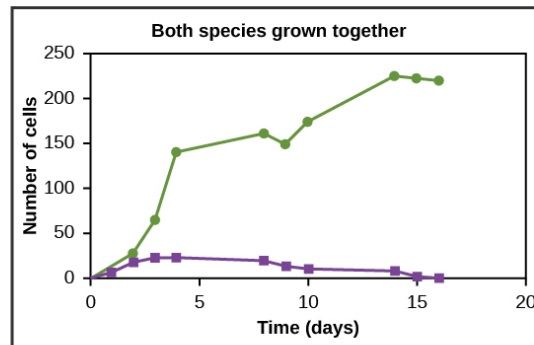




(a)



(b)



(c)

*Paramecium aurelia* and *Paramecium caudatum* grow well individually, but when they compete for the same resources, the *P. aurelia* outcompetes the *P. caudatum*.

This exclusion may be avoided if a population evolves to make use of a different resource, a different area of the habitat, or feeds during a different time of day, called resource partitioning. The two organisms are then said to occupy different microniches. These organisms coexist by minimizing direct competition.

## Symbiosis

Symbiotic relationships, or **symbioses** (plural), are close interactions between individuals of different species over an extended period of time which impact the abundance and distribution of the associating populations. Most scientists accept this definition, but some restrict the term to only

those species that are mutualistic, where both individuals benefit from the interaction. In this discussion, the broader definition will be used.

## Commensalism

A **commensal** relationship occurs when one species benefits from the close, prolonged interaction, while the other neither benefits nor is harmed. Birds nesting in trees provide an example of a commensal relationship ([link](#)). The tree is not harmed by the presence of the nest among its branches. The nests are light and produce little strain on the structural integrity of the branch, and most of the leaves, which the tree uses to get energy by photosynthesis, are above the nest so they are unaffected. The bird, on the other hand, benefits greatly. If the bird had to nest in the open, its eggs and young would be vulnerable to predators. Another example of a commensal relationship is the clown fish and the sea anemone. The sea anemone is not harmed by the fish, and the fish benefits with protection from predators who would be stung upon nearing the sea anemone.



The southern masked-weaver bird is starting to make a nest in a tree in Zambezi Valley,

Zambia. This is an example of a commensal relationship, in which one species (the bird) benefits, while the other (the tree) neither benefits nor is harmed. (credit: “Hanay”/Wikimedia Commons)

## Mutualism

A second type of symbiotic relationship is called **mutualism**, where two species benefit from their interaction. Some scientists believe that these are the only true examples of symbiosis. For example, termites have a mutualistic relationship with protozoa that live in the insect’s gut ([link](#)**a**). The termite benefits from the ability of bacterial symbionts within the protozoa to digest cellulose. The termite itself cannot do this, and without the protozoa, it would not be able to obtain energy from its food (cellulose from the wood it chews and eats). The protozoa and the bacterial symbionts benefit by having a protective environment and a constant supply of food from the wood chewing actions of the termite. Lichens have a mutualistic relationship between fungus and photosynthetic algae or bacteria ([link](#)**b**). As these symbionts grow together, the glucose produced by the algae provides nourishment for both organisms, whereas the physical structure of the lichen protects the algae from the elements and makes certain nutrients in the atmosphere more available to the algae.



(a) Termites form a mutualistic relationship with symbiotic protozoa in their guts, which allow both organisms to obtain energy from the cellulose the termite consumes. (b) Lichen is a fungus that has symbiotic photosynthetic algae living inside its cells.

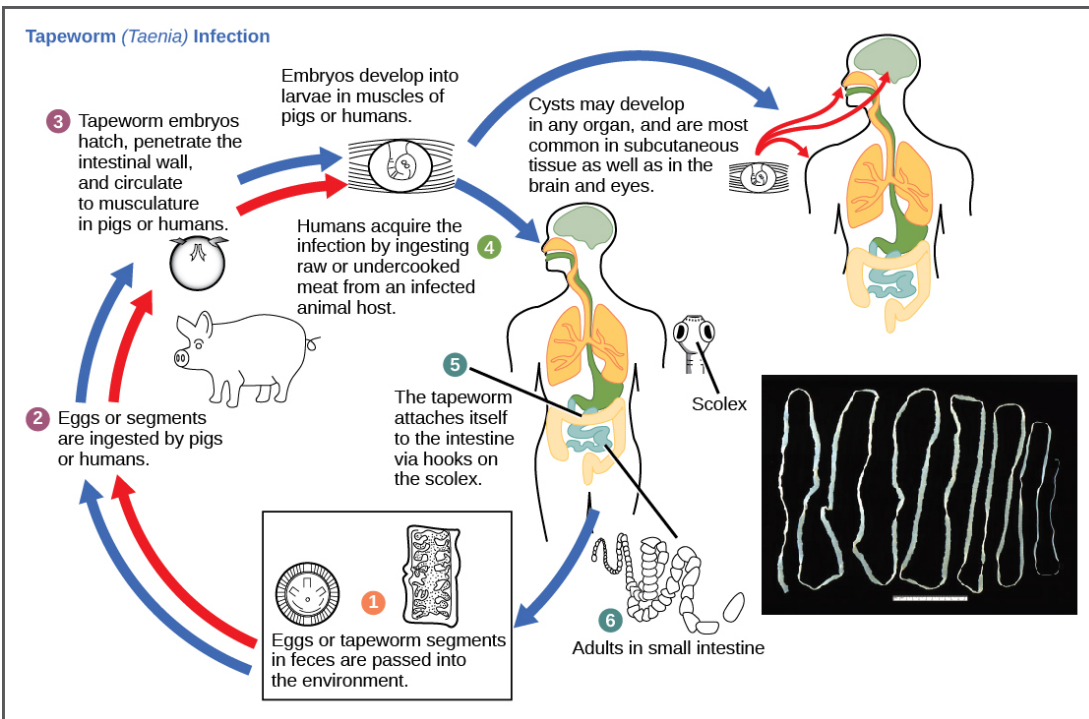
(credit a: modification of work by Scott Bauer, USDA; credit b: modification of work by Cory Zanker)

## Parasitism

A **parasite** is an organism that lives in or on another living organism and derives nutrients from it. In this relationship, the parasite benefits, but the organism being fed upon, the **host** is harmed. The host is usually weakened by the parasite as it siphons resources the host would normally use to maintain itself. The parasite, however, is unlikely to kill the host, especially not quickly, because this would allow no time for the organism to complete its reproductive cycle by spreading to another host.

The reproductive cycles of parasites are often very complex, sometimes requiring more than one host species. A tapeworm is a parasite that causes disease in humans when contaminated, undercooked meat such as pork, fish, or beef is consumed ([link](#)). The tapeworm can live inside the intestine of the host for several years, benefiting from the food the host is bringing into its gut by eating, and may grow to be over 50 ft long by adding segments. The parasite moves from species to species in a cycle,

making two hosts necessary to complete its life cycle. Another common parasite is *Plasmodium falciparum*, the protozoan cause of malaria, a significant disease in many parts of the world. Living in human liver and red blood cells, the organism reproduces asexually in the gut of blood-feeding mosquitoes to complete its life cycle. Thus malaria is spread from human to human by mosquitoes, one of many arthropod-borne infectious diseases.



This diagram shows the life cycle of a pork tapeworm (*Taenia solium*), a human worm parasite. (credit: modification of work by CDC)

## Characteristics of Communities

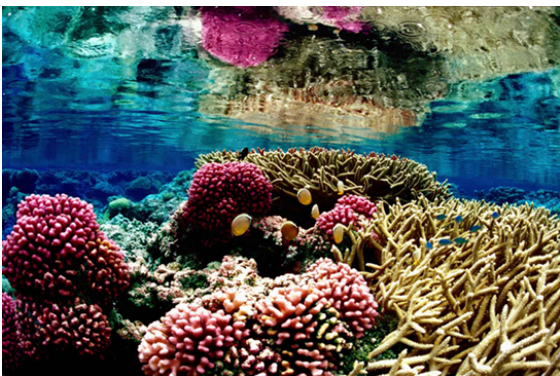
Communities are complex entities that can be characterized by their structure (the types and numbers of species present) and dynamics (how

communities change over time). Understanding community structure and dynamics enables community ecologists to manage ecosystems more effectively.

## Foundation Species

**Foundation species** are considered the “base” or “bedrock” of a community, having the greatest influence on its overall structure. They are usually the primary producers: organisms that bring most of the energy into the community. Kelp, brown algae, is a foundation species, forming the basis of the kelp forests off the coast of California.

Foundation species may physically modify the environment to produce and maintain habitats that benefit the other organisms that use them. An example is the photosynthetic corals of the coral reef ([link](#)). Corals themselves are not photosynthetic, but harbor symbionts within their body tissues (dinoflagellates called zooxanthellae) that perform photosynthesis; this is another example of a mutualism. The exoskeletons of living and dead coral make up most of the reef structure, which protects many other species from waves and ocean currents.

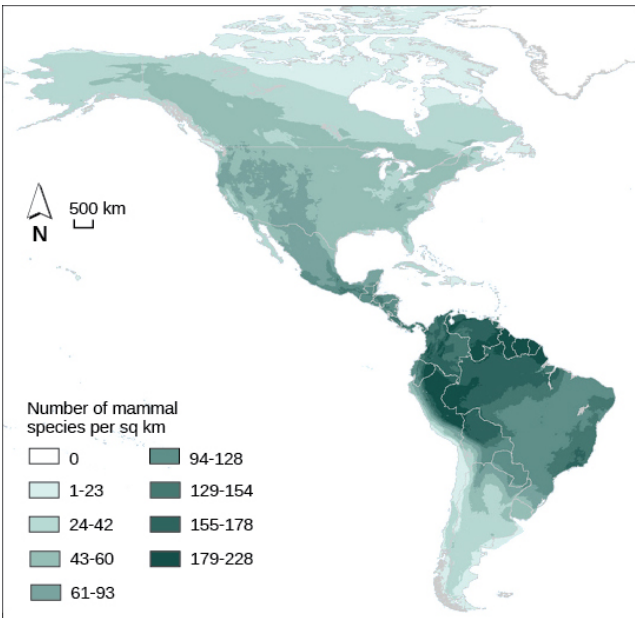


Coral is the foundation species of coral reef ecosystems. (credit: Jim E. Maragos, USFWS)



## **Biodiversity, Species Richness, and Relative Species Abundance**

Biodiversity describes a community's biological complexity: it is measured by the number of different species (species richness) in a particular area and their relative abundance (species evenness). The area in question could be a habitat, a biome, or the entire biosphere. **Species richness** is the term that is used to describe the number of species living in a habitat or biome. Species richness varies across the globe ([link](#)). One factor in determining species richness is latitude, with the greatest species richness occurring in ecosystems near the equator, which often have warmer temperatures, large amounts of rainfall, and low seasonality. The lowest species richness occurs near the poles, which are much colder, drier, and thus less conducive to life in Geologic time (time since glaciations). The predictability of climate or productivity is also an important factor. Other factors influence species richness as well. For example, the study of **island biogeography** attempts to explain the relatively high species richness found in certain isolated island chains, including the Galápagos Islands that inspired the young Darwin. **Relative species abundance** is the number of individuals in a species relative to the total number of individuals in all species within a habitat, ecosystem, or biome. Foundation species often have the highest relative abundance of species.



The greatest species richness for mammals in North and South America is associated with the equatorial latitudes. (credit: modification of work by NASA, CIESIN, Columbia University)

## Keystone Species

A **keystone species** is one whose presence is key to maintaining biodiversity within an ecosystem and to upholding an ecological community's structure. The intertidal sea star, *Pisaster ochraceus*, of the northwestern United States is a keystone species ([link](#)). Studies have shown that when this organism is removed from communities, populations of their natural prey (mussels) increase, completely altering the species composition and reducing biodiversity. Another keystone species is the banded tetra, a fish in tropical streams, which supplies nearly all of the phosphorus, a necessary inorganic nutrient, to the rest of the community. If



these fish were to become extinct, the community would be greatly affected.



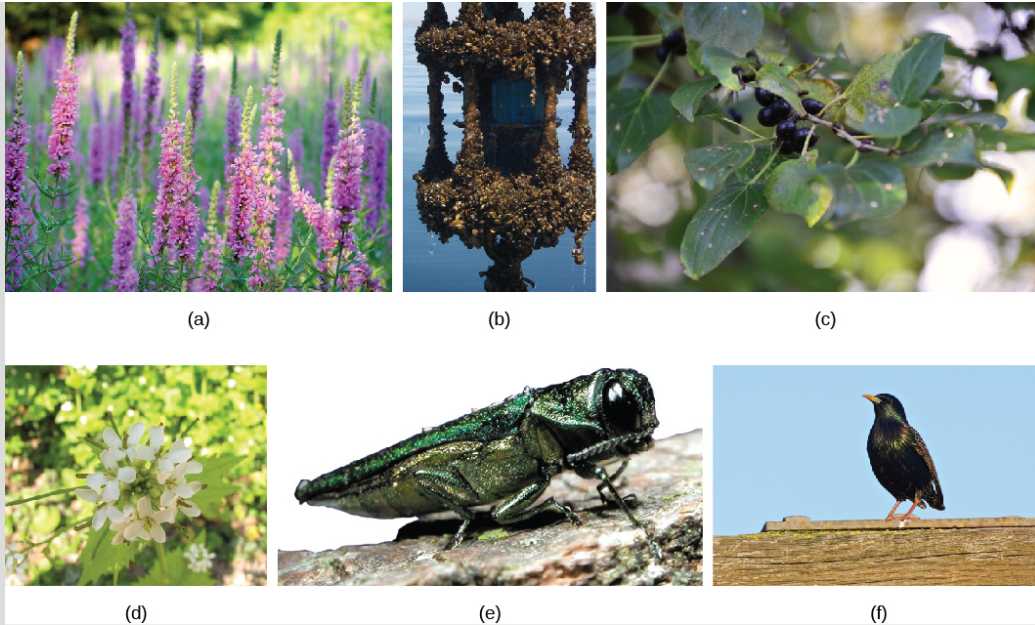
The *Pisaster ochraceus* sea star is a keystone species.  
(credit: Jerry Kirkhart)

**Note:**

**Everyday Connection**

**Invasive Species**

Invasive species are non-native organisms that, when introduced to an area out of their native range, threaten the ecosystem balance of that habitat. Many such species exist in the United States, as shown in [\[link\]](#). Whether enjoying a forest hike, taking a summer boat trip, or simply walking down an urban street, you have likely encountered an invasive species.



In the United States, invasive species like (a) purple loosestrife (*Lythrum salicaria*) and the (b) zebra mussel (*Dreissena polymorpha*) threaten certain aquatic ecosystems. Some forests are threatened by the spread of (c) common buckthorn (*Rhamnus cathartica*), (d) garlic mustard (*Alliaria petiolata*), and (e) the emerald ash borer (*Agrilus planipennis*). The (f) European starling (*Sturnus vulgaris*) may compete with native bird species for nest holes. (credit a: modification of work by Liz West; credit b: modification of work by M. McCormick, NOAA; credit c: modification of work by E. Dronkert; credit d: modification of work by Dan Davison; credit e: modification of work by USDA; credit f: modification of work by Don DeBold)

One of the many recent proliferations of an invasive species concerns the growth of Asian carp populations. Asian carp were introduced to the United States in the 1970s by fisheries and sewage treatment facilities that used the fish's excellent filter feeding capabilities to clean their ponds of excess plankton. Some of the fish escaped, however, and by the 1980s they had colonized many waterways of the Mississippi River basin, including the Illinois and Missouri Rivers.

Voracious eaters and rapid reproducers, Asian carp may outcompete native species for food, potentially leading to their extinction. For example, black carp are voracious eaters of native mussels and snails, limiting this food source for native fish species. Silver carp eat plankton that native mussels and snails feed on, reducing this food source by a different alteration of the food web. In some areas of the Mississippi River, Asian carp species have become the most predominant, effectively outcompeting native fishes for habitat. In some parts of the Illinois River, Asian carp constitute 95 percent of the community's biomass. Although edible, the fish is bony and not a desired food in the United States. Moreover, their presence threatens the native fish and fisheries of the Great Lakes, which are important to local economies and recreational anglers. Asian carp have even injured humans. The fish, frightened by the sound of approaching motorboats, thrust themselves into the air, often landing in the boat or directly hitting the boaters.

The Great Lakes and their prized salmon and lake trout fisheries are also being threatened by these invasive fish. Asian carp have already colonized rivers and canals that lead into Lake Michigan. One infested waterway of particular importance is the Chicago Sanitary and Ship Channel, the major supply waterway linking the Great Lakes to the Mississippi River. To prevent the Asian carp from leaving the canal, a series of electric barriers have been successfully used to discourage their migration; however, the threat is significant enough that several states and Canada have sued to have the Chicago channel permanently cut off from Lake Michigan. Local and national politicians have weighed in on how to solve the problem, but no one knows whether the Asian carp will ultimately be considered a nuisance, like other invasive species such as the water hyacinth and zebra mussel, or whether it will be the destroyer of the largest freshwater fishery of the world.

The issues associated with Asian carp show how population and community ecology, fisheries management, and politics intersect on issues of vital importance to the human food supply and economy. Socio-political issues like this make extensive use of the sciences of population ecology (the study of members of a particular species occupying a particular area known as a habitat) and community ecology (the study of the interaction of all species within a habitat).

## Community Dynamics

Community dynamics are the changes in community structure and composition over time. Sometimes these changes are induced by **environmental disturbances** such as volcanoes, earthquakes, storms, fires, and climate change. Communities with a stable structure are said to be at equilibrium. Following a disturbance, the community may or may not return to the equilibrium state.

Succession describes the sequential appearance and disappearance of species in a community over time. In **primary succession**, newly exposed or newly formed land is colonized by living things; in **secondary succession**, part of an ecosystem is disturbed and remnants of the previous community remain.

### Primary Succession and Pioneer Species

Primary succession occurs when new land is formed or rock is exposed: for example, following the eruption of volcanoes, such as those on the Big Island of Hawaii. As lava flows into the ocean, new land is continually being formed. On the Big Island, approximately 32 acres of land is added each year. First, weathering and other natural forces break down the substrate enough for the establishment of certain hearty plants and lichens with few soil requirements, known as **pioneer species** ([link](#)). These species help to further break down the mineral rich lava into soil where other, less hardy species will grow and eventually replace the pioneer species. In addition, as these early species grow and die, they add to an ever-growing layer of decomposing organic material and contribute to soil formation. Over time the area will reach an equilibrium state, with a set of organisms quite different from the pioneer species.



During primary succession in lava on Maui, Hawaii, succulent plants are the pioneer species. (credit: Forest and Kim Starr)

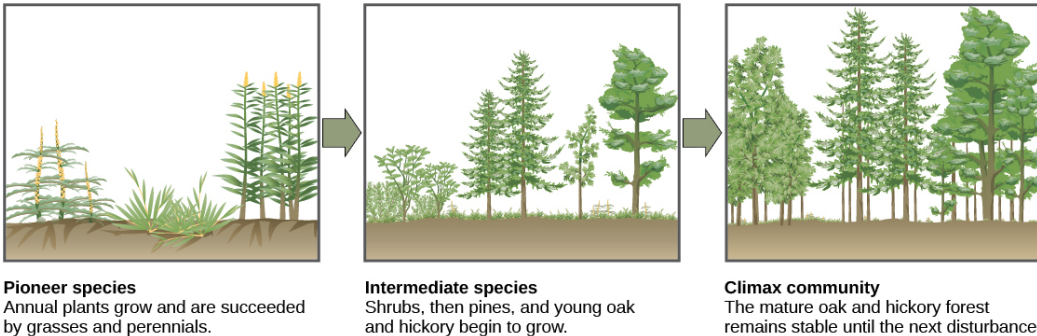
## **Secondary succession**

A classic example of secondary succession occurs in oak and hickory forests cleared by wildfire ([link](#)). Wildfires will burn most vegetation and kill those animals unable to flee the area. Their nutrients, however, are returned to the ground in the form of ash. Thus, even when areas are devoid of life due to severe fires, the area will soon be ready for new life to take hold.

Before the fire, the vegetation was dominated by tall trees with access to the major plant energy resource: sunlight. Their height gave them access to sunlight while also shading the ground and other low-lying species. After the fire, though, these trees are no longer dominant. Thus, the first plants to grow back are usually annual plants followed within a few years by quickly growing and spreading grasses and other pioneer species. Due to, at least in part, changes in the environment brought on by the growth of the grasses and other species, over many years, shrubs will emerge along with small pine, oak, and hickory trees. These organisms are called intermediate

species. Eventually, over 150 years, the forest will reach its equilibrium point where species composition is no longer changing and resembles the community before the fire. This equilibrium state is referred to as the **climax community**, which will remain stable until the next disturbance.

Secondary Succession of an Oak and Hickory Forest



Secondary succession is shown in an oak and hickory forest after a forest fire.

## Section Summary

Communities include all the different species living in a given area. The variety of these species is called species richness. Many organisms have developed defenses against predation and herbivory, including mechanical defenses, warning coloration, and mimicry, as a result of evolution and the interaction with other members of the community. Two species cannot exist in the same habitat competing directly for the same resources. Species may form symbiotic relationships such as commensalism or mutualism.

Community structure is described by its foundation and keystone species. Communities respond to environmental disturbances by succession (the predictable appearance of different types of plant species) until a stable community structure is established.

## Review Questions

**Exercise:**

**Problem:**

The first species to live on new land, such as that formed from volcanic lava, are called \_\_\_\_\_.

- a. climax community
- b. keystone species
- c. foundation species
- d. pioneer species

---

**Solution:**

D

**Exercise:**

**Problem:**

Which type of mimicry involves multiple species with similar warning coloration that are all toxic to predators?

- a. Batesian mimicry
- b. Müllerian mimicry
- c. Emsleyan/Mertensian mimicry
- d. Mertensian mimicry

---

**Solution:**

B

**Exercise:**

**Problem:**

A symbiotic relationship where both of the coexisting species benefit from the interaction is called \_\_\_\_\_.



- a. commensalism
- b. parasitism
- c. mutualism
- d. communism

---

**Solution:**

C

## Free Response

**Exercise:**

**Problem:**

Describe the competitive exclusion principle and its effects on competing species.

---

**Solution:**

The competitive exclusion principle states that no two species competing for the same resources at the same time and place can coexist over time. Thus, one of the competing species will eventually dominate. On the other hand, if the species evolve such that they use resources from different parts of the habitat or at different times of day, the two species can exist together indefinitely.

## Glossary

**aposematic coloration**

warning coloration used as a defensive mechanism against predation

**Batesian mimicry**

type of mimicry where a non-harmful species takes on the warning colorations of a harmful one



camouflage

avoid detection by blending in with the background.

climax community

final stage of succession, where a stable community is formed by a characteristic assortment of plant and animal species

commensalism

relationship between species wherein one species benefits from the close, prolonged interaction, while the other species neither benefits nor is harmed

competitive exclusion principle

no two species within a habitat can coexist when they compete for the same resources at the same place and time

Emsleyan/Mertensian mimicry

type of mimicry where a harmful species resembles a less harmful one

environmental disturbance

change in the environment caused by natural disasters or human activities

foundation species

species which often forms the major structural portion of the habitat

host

organism a parasite lives on

island biogeography

study of life on island chains and how their geography interacts with the diversity of species found there

keystone species

species whose presence is key to maintaining biodiversity in an ecosystem and to upholding an ecological community's structure

Müllerian mimicry

type of mimicry where species share warning coloration and all are harmful to predators

mutualism

symbiotic relationship between two species where both species benefit

parasite

organism that uses resources from another species, the host

pioneer species

first species to appear in primary and secondary succession

primary succession

succession on land that previously has had no life

relative species abundance

absolute population size of a particular species relative to the population sizes of other species within the community

secondary succession

succession in response to environmental disturbances that move a community away from its equilibrium

species richness

number of different species in a community

symbiosis

close interaction between individuals of different species over an extended period of time that impacts the abundance and distribution of the associating populations

## Behavioral Biology: Proximate and Ultimate Causes of Behavior

By the end of this section, you will be able to:

- Compare innate and learned behavior
- Discuss how movement and migration behaviors are a result of natural selection
- Discuss the different ways members of a population communicate with each other
- Give examples of how species use energy for mating displays and other courtship behaviors
- Differentiate between various mating systems
- Describe different ways that species learn

**Behavior** is the change in activity of an organism in response to a stimulus. **Behavioral biology** is the study of the biological and evolutionary bases for such changes. The idea that behaviors evolved as a result of the pressures of natural selection is not new. Animal behavior has been studied for decades, by biologists in the science of **ethology**, by psychologists in the science of comparative psychology, and by scientists of many disciplines in the study of neurobiology. Although there is overlap between these disciplines, scientists in these behavioral fields take different approaches. Comparative psychology is an extension of work done in human and behavioral psychology. Ethology is an extension of genetics, evolution, anatomy, physiology, and other biological disciplines. Still, one cannot study behavioral biology without touching on both comparative psychology and ethology.

One goal of behavioral biology is to dissect out the **innate behaviors**, which have a strong genetic component and are largely independent of environmental influences, from the **learned behaviors**, which result from environmental conditioning. Innate behavior, or instinct, is important because there is no risk of an incorrect behavior being learned. They are “hard wired” into the system. On the other hand, learned behaviors, although riskier, are flexible, dynamic, and can be altered according to changes in the environment.

### Innate Behaviors: Movement and Migration

Innate or instinctual behaviors rely on response to stimuli. The simplest example of this is a **reflex action**, an involuntary and rapid response to stimulus. To test the “knee-jerk” reflex, a doctor taps the patellar tendon below the kneecap with a rubber hammer. The stimulation of the nerves there leads to the reflex of extending the leg at the knee. This is similar to the reaction of someone who touches a hot stove and instinctually pulls his or her hand away. Even humans, with our great capacity to learn, still exhibit a variety of innate behaviors.

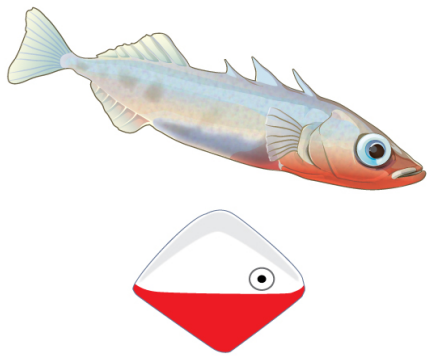
## **Kinesis and Taxis**

Another activity or movement of innate behavior is **kinesis**, or the undirected movement in response to a stimulus. Orthokinesis is the increased or decreased speed of movement of an organism in response to a stimulus. Woodlice, for example, increase their speed of movement when exposed to high or low temperatures. This movement, although random, increases the probability that the insect spends less time in the unfavorable environment. Another example is klinokinesis, an increase in turning behaviors. It is exhibited by bacteria such as *E. coli* which, in association with orthokinesis, helps the organisms randomly find a more hospitable environment.

A similar, but more directed version of kinesis is **taxis**: the directed movement towards or away from a stimulus. This movement can be in response to light (phototaxis), chemical signals (chemotaxis), or gravity (geotaxis) and can be directed toward (positive) or away (negative) from the source of the stimulus. An example of a positive chemotaxis is exhibited by the unicellular protozoan *Tetrahymena thermophila*. This organism swims using its cilia, at times moving in a straight line, and at other times making turns. The attracting chemotactic agent alters the frequency of turning as the organism moves directly toward the source, following the increasing concentration gradient.

## **Fixed Action Patterns**

A **fixed action pattern** is a series of movements elicited by a stimulus such that even when the stimulus is removed, the pattern goes on to completion. An example of such a behavior occurs in the three-spined stickleback, a small freshwater fish ([link](#)). Males of this species develop a red belly during breeding season and show instinctual aggressiveness to other males during this time. In laboratory experiments, researchers exposed such fish to objects that in no way resemble a fish in their shape, but which were painted red on their lower halves. The male sticklebacks responded aggressively to the objects just as if they were real male sticklebacks.



Male three-spined stickleback fish exhibit a fixed action pattern. During mating season, the males, which develop a bright red belly, react strongly to red-bottomed objects that in no way resemble fish.

## Migration

**Migration** is the long-range seasonal movement of animals. It is an evolved, adapted response to variation in resource availability, and it is a common phenomenon found in all major groups of animals. Birds fly south for the winter to get to warmer climates with sufficient food, and salmon migrate to their spawning grounds. The popular 2005 documentary *March of the Penguins* followed the 62-mile migration of emperor penguins through Antarctica to bring food back to their breeding site and to their young. Wildebeests ([link](#)) migrate over 1800 miles each year in search of new grasslands.



Wildebeests migrate in a clockwise fashion over 1800 miles each year in search of rain-ripened grass. (credit: Eric Inafuku)

Although migration is thought of as innate behavior, only some migrating species always migrate (obligate migration). Animals that exhibit facultative migration can choose to migrate or not. Additionally, in some animals, only a portion of the population migrates, whereas the rest does not migrate (incomplete migration). For example, owls that live in the tundra may migrate in years when their food source, small rodents, is relatively scarce, but not migrate during the years when rodents are plentiful.

## Foraging

**Foraging** is the act of searching for and exploiting food resources. Feeding behaviors that maximize energy gain and minimize energy expenditure are called optimal foraging behaviors, and these are favored by natural selection. The painted stork, for example, uses its long beak to search the bottom of a freshwater marshland for crabs and other food ([link](#)).



The painted stork uses its long beak to forage. (credit: J.M. Garg)

## Innate Behaviors: Living in Groups

Not all animals live in groups, but even those that live relatively solitary lives, with the exception of those that can reproduce asexually, must mate. Mating usually involves one animal signaling another so as to communicate the desire to mate. There are several types of energy-intensive behaviors or displays associated with mating, called mating rituals. Other behaviors found in populations that live in groups are described in terms of which animal benefits from the behavior. In selfish behavior, only the animal in question benefits; in altruistic behavior, one animal's actions benefit another animal; cooperative behavior describes when both animals benefit. All of

these behaviors involve some sort of communication between population members.

## **Communication within a Species**

Animals communicate with each other using stimuli known as **signals**. An example of this is seen in the three-spined stickleback, where the visual signal of a red region in the lower half of a fish signals males to become aggressive and signals females to mate. Other signals are chemical (pheromones), aural (sound), visual (courtship and aggressive displays), or tactile (touch). These types of communication may be instinctual or learned or a combination of both. These are not the same as the communication we associate with language, which has been observed only in humans and perhaps in some species of primates and cetaceans.

A pheromone is a secreted chemical signal used to obtain a response from another individual of the same species. The purpose of pheromones is to elicit a specific behavior from the receiving individual. Pheromones are especially common among social insects, but they are used by many species to attract the opposite sex, to sound alarms, to mark food trails, and to elicit other, more complex behaviors. Even humans are thought to respond to certain pheromones called axillary steroids. These chemicals influence human perception of other people, and in one study were responsible for a group of women synchronizing their menstrual cycles. The role of pheromones in human-to-human communication is still somewhat controversial and continues to be researched.

Songs are an example of an aural signal, one that needs to be heard by the recipient. Perhaps the best known of these are songs of birds, which identify the species and are used to attract mates. Other well-known songs are those of whales, which are of such low frequency that they can travel long distances underwater. Dolphins communicate with each other using a wide variety of vocalizations. Male crickets make chirping sounds using a specialized organ to attract a mate, repel other males, and to announce a successful mating.



**Courtship displays** are a series of ritualized visual behaviors (signals) designed to attract and convince a member of the opposite sex to mate. These displays are ubiquitous in the animal kingdom. Often these displays involve a series of steps, including an initial display by one member followed by a response from the other. If at any point, the display is performed incorrectly or a proper response is not given, the mating ritual is abandoned and the mating attempt will be unsuccessful. The mating display of the common stork is shown in [\[link\]](#).

**Aggressive displays** are also common in the animal kingdom. An example is when a dog bares its teeth when it wants another dog to back down. Presumably, these displays communicate not only the willingness of the animal to fight, but also its fighting ability. Although these displays do signal aggression on the part of the sender, it is thought that these displays are actually a mechanism to reduce the amount of actual fighting that occurs between members of the same species: they allow individuals to assess the fighting ability of their opponent and thus decide whether it is “worth the fight.” The testing of certain hypotheses using game theory has led to the conclusion that some of these displays may overstate an animal’s actual fighting ability and are used to “bluff” the opponent. This type of interaction, even if “dishonest,” would be favored by natural selection if it is successful more times than not.



This stork’s courtship display is designed to attract potential

mates. (credit: Linda  
“jinterwas”/Flickr)

**Distraction displays** are seen in birds and some fish. They are designed to attract a predator away from the nest that contains their young. This is an example of an altruistic behavior: it benefits the young more than the individual performing the display, which is putting itself at risk by doing so.

Many animals, especially primates, communicate with other members in the group through touch. Activities such as grooming, touching the shoulder or root of the tail, embracing, lip contact, and greeting ceremonies have all been observed in the Indian langur, an Old World monkey. Similar behaviors are found in other primates, especially in the great apes.

**Note:**

Link to Learning



The killdeer bird distracts predators from its eggs by faking a broken wing display in this video taken in Boise, Idaho.

[https://www.openstaxcollege.org/l/killdeer\\_bird](https://www.openstaxcollege.org/l/killdeer_bird)

## Altruistic Behaviors

Behaviors that lower the fitness of the individual but increase the fitness of another individual are termed altruistic. Examples of such behaviors are seen widely across the animal kingdom. Social insects such as worker bees have no ability to reproduce, yet they maintain the queen so she can

populate the hive with her offspring. Meerkats keep a sentry standing guard to warn the rest of the colony about intruders, even though the sentry is putting itself at risk. Wolves and wild dogs bring meat to pack members not present during a hunt. Lemurs take care of infants unrelated to them. Although on the surface, these behaviors appear to be altruistic, it may not be so simple.

There has been much discussion over why altruistic behaviors exist. Do these behaviors lead to overall evolutionary advantages for their species? Do they help the altruistic individual pass on its own genes? And what about such activities between unrelated individuals? One explanation for altruistic-type behaviors is found in the genetics of natural selection. In the 1976 book, *The Selfish Gene*, scientist Richard Dawkins attempted to explain many seemingly altruistic behaviors from the viewpoint of the gene itself. Although a gene obviously cannot be selfish in the human sense, it may appear that way if the sacrifice of an individual benefits related individuals that share genes that are identical by descent (present in relatives because of common lineage). Mammal parents make this sacrifice to take care of their offspring. Emperor penguins migrate miles in harsh conditions to bring food back for their young. Selfish gene theory has been controversial over the years and is still discussed among scientists in related fields.

Even less-related individuals, those with less genetic identity than that shared by parent and offspring, benefit from seemingly altruistic behavior. The activities of social insects such as bees, wasps, ants, and termites are good examples. Sterile workers in these societies take care of the queen because they are closely related to it, and as the queen has offspring, she is passing on genes from the workers indirectly. Thus, it is of fitness benefit for the worker to maintain the queen without having any direct chance of passing on its genes due to its sterility. The lowering of individual fitness to enhance the reproductive fitness of a relative and thus one's inclusive fitness evolves through **kin selection**. This phenomenon can explain many superficially altruistic behaviors seen in animals. However, these behaviors may not be truly defined as altruism in these cases because the actor is actually increasing its own fitness either directly (through its own offspring)

or indirectly (through the inclusive fitness it gains through relatives that share genes with it).

Unrelated individuals may also act altruistically to each other, and this seems to defy the “selfish gene” explanation. An example of this observed in many monkey species where a monkey will present its back to an unrelated monkey to have that individual pick the parasites from its fur. After a certain amount of time, the roles are reversed and the first monkey now grooms the second monkey. Thus, there is reciprocity in the behavior. Both benefit from the interaction and their fitness is raised more than if neither cooperated nor if one cooperated and the other did not cooperate. This behavior is still not necessarily altruism, as the “giving” behavior of the actor is based on the expectation that it will be the “receiver” of the behavior in the future, termed reciprocal altruism. Reciprocal altruism requires that individuals repeatedly encounter each other, often the result of living in the same social group, and that cheaters (those that never “give back”) are punished.

Evolutionary game theory, a modification of classical game theory in mathematics, has shown that many of these so-called “altruistic behaviors” are not altruistic at all. The definition of “pure” altruism, based on human behavior, is an action that benefits another without any direct benefit to oneself. Most of the behaviors previously described do not seem to satisfy this definition, and game theorists are good at finding “selfish” components in them. Others have argued that the terms “selfish” and “altruistic” should be dropped completely when discussing animal behavior, as they describe human behavior and may not be directly applicable to instinctual animal activity. What is clear, though, is that heritable behaviors that improve the chances of passing on one’s genes or a portion of one’s genes are favored by natural selection and will be retained in future generations as long as those behaviors convey a fitness advantage. These instinctual behaviors may then be applied, in special circumstances, to other species, as long as it doesn’t lower the animal’s fitness.

## **Finding Sex Partners**

Not all animals reproduce sexually, but many that do have the same challenge: they need to find a suitable mate and often have to compete with other individuals to obtain one. Significant energy is spent in the process of locating, attracting, and mating with the sex partner. Two types of selection occur during this process and can lead to traits that are important to reproduction called secondary sexual characteristics: **intersexual selection**, the choosing of a mate where individuals of one sex choose mates of the other sex, and **intrasexual selection**, the competition for mates between species members of the same sex. Intersexual selection is often complex because choosing a mate may be based on a variety of visual, aural, tactile, and chemical cues. An example of intersexual selection is when female peacocks choose to mate with the male with the brightest plumage. This type of selection often leads to traits in the chosen sex that do not enhance survival, but are those traits most attractive to the opposite sex (often at the expense of survival). Intrasexual selection involves mating displays and aggressive mating rituals such as rams butting heads—the winner of these battles is the one that is able to mate. Many of these rituals use up considerable energy but result in the selection of the healthiest, strongest, and/or most dominant individuals for mating. Three general mating systems, all involving innate as opposed to learned behaviors, are seen in animal populations: monogamous, polygynous, and polyandrous.

**Note:**

Link to Learning

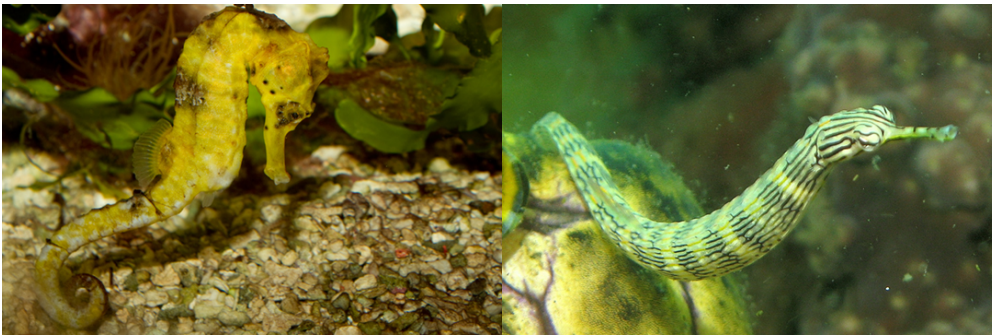


Visit this [website](#) for informative videos on sexual selection.

In **monogamous** systems, one male and one female are paired for at least one breeding season. In some animals, such as the gray wolf, these associations can last much longer, even a lifetime. Several explanations have been proposed for this type of mating system. The “mate-guarding hypothesis” states that males stay with the female to prevent other males from mating with her. This behavior is advantageous in such situations where mates are scarce and difficult to find. Another explanation is the “male-assistance hypothesis,” where males that remain with a female to help guard and rear their young will have more and healthier offspring. Monogamy is observed in many bird populations where, in addition to the parental care from the female, the male is also a major provider of parental care for the chicks. A third explanation for the evolutionary advantages of monogamy is the “female-enforcement hypothesis.” In this scenario, the female ensures that the male does not have other offspring that might compete with her own, so she actively interferes with the male’s signaling to attract other mates.

**Polygynous** mating refers to one male mating with multiple females. In these situations, the female must be responsible for most of the parental care as the single male is not capable of providing care to that many offspring. In resource-based polygyny, males compete for territories with the best resources, and then mate with females that enter the territory, drawn to its resource richness. The female benefits by mating with a dominant, genetically fit male; however, it is at the cost of having no male help in caring for the offspring. An example is seen in the yellow-rumped honeyguide, a bird whose males defend beehives because the females feed on their wax. As the females approach, the male defending the nest will mate with them. Harem mating structures are a type of polygynous system where certain males dominate mating while controlling a territory with resources. Elephant seals, where the alpha male dominates the mating within the group are an example. A third type of polygyny is a lek system. Here there is a communal courting area where several males perform elaborate displays for females, and the females choose their mate from this group. This behavior is observed in several bird species including the sage grouse and the prairie chicken.

In **polyandrous** mating systems, one female mates with many males. These types of systems are much rarer than monogamous and polygynous mating systems. In pipefishes and seahorses, males receive the eggs from the female, fertilize them, protect them within a pouch, and give birth to the offspring ([link](#)). Therefore, the female is able to provide eggs to several males without the burden of carrying the fertilized eggs.



Polyandrous mating, in which one female mates with many males, occurs in the (a) seahorse and the (b) pipefish. (credit a: modification of work by Brian Gratwicke; credit b: modification of work by Stephen Childs)

## Simple Learned Behaviors

The majority of the behaviors previously discussed were innate or at least have an innate component (variations on the innate behaviors may be learned). They are inherited and the behaviors do not change in response to signals from the environment. Conversely, learned behaviors, even though they may have instinctive components, allow an organism to adapt to changes in the environment and are modified by previous experiences. Simple learned behaviors include habituation and imprinting—both are important to the maturation process of young animals.

## Habituation

**Habituation** is a simple form of learning in which an animal stops responding to a stimulus after a period of repeated exposure. This is a form of non-associative learning, as the stimulus is not associated with any punishment or reward. Prairie dogs typically sound an alarm call when threatened by a predator, but they become habituated to the sound of human footsteps when no harm is associated with this sound, therefore, they no longer respond to them with an alarm call. In this example, habituation is specific to the sound of human footsteps, as the animals still respond to the sounds of potential predators.

## Imprinting

**Imprinting** is a type of learning that occurs at a particular age or a life stage that is rapid and independent of the species involved. Hatchling ducks recognize the first adult they see, their mother, and make a bond with her. A familiar sight is ducklings walking or swimming after their mothers ([\[link\]](#)). This is another type of non-associative learning, but is very important in the maturation process of these animals as it encourages them to stay near their mother so they will be protected, greatly increasing their chances of survival. However, if newborn ducks see a human before they see their mother, they will imprint on the human and follow it in just the same manner as they would follow their real mother.





The attachment of ducklings to their mother is an example of imprinting. (credit: modification of work by Mark Harkin)

**Note:**

Link to Learning



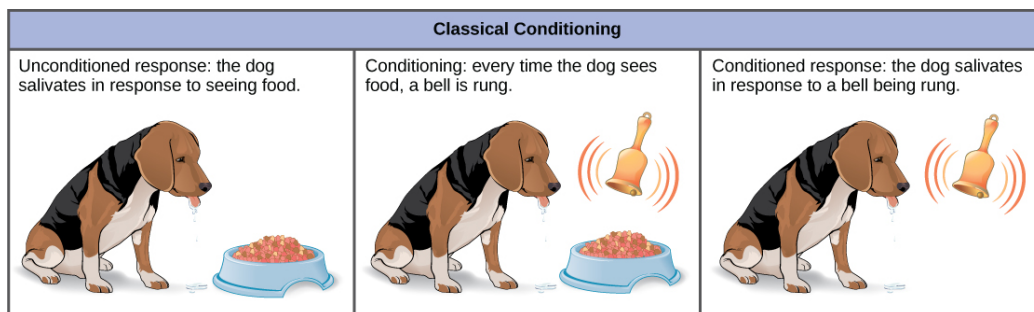
The International Crane Foundation has helped raise the world's population of whooping cranes from 21 individuals to about 600. Imprinting hatchlings has been a key to success: biologists wear full crane costumes so the birds never “see” humans. Watch this video to learn more. [https://www.openstaxcollege.org/l/whooping\\_crane](https://www.openstaxcollege.org/l/whooping_crane)

## Conditioned Behavior

**Conditioned behaviors** are types of associative learning, where a stimulus becomes associated with a consequence. During operant conditioning, the behavioral response is modified by its consequences, with regards to its form, strength, or frequency.

### Classical Conditioning

In **classical conditioning**, a response called the conditioned response is associated with a stimulus that it had previously not been associated with, the conditioned stimulus. The response to the original, unconditioned stimulus is called the unconditioned response. The most cited example of classical conditioning is Ivan Pavlov's experiments with dogs ([link](#)). In Pavlov's experiments, the unconditioned response was the salivation of dogs in response to the unconditioned stimulus of seeing or smelling their food. The conditioning stimulus that researchers associated with the unconditioned response was the ringing of a bell. During conditioning, every time the animal was given food, the bell was rung. This was repeated during several trials. After some time, the dog learned to associate the ringing of the bell with food and to respond by salivating. After the conditioning period was finished, the dog would respond by salivating when the bell was rung, even when the unconditioned stimulus, the food, was absent. Thus, the ringing of the bell became the conditioned stimulus and the salivation became the conditioned response. Although it is thought by some scientists that the unconditioned and conditioned responses are identical, even Pavlov discovered that the saliva in the conditioned dogs had characteristic differences when compared to the unconditioned dog.



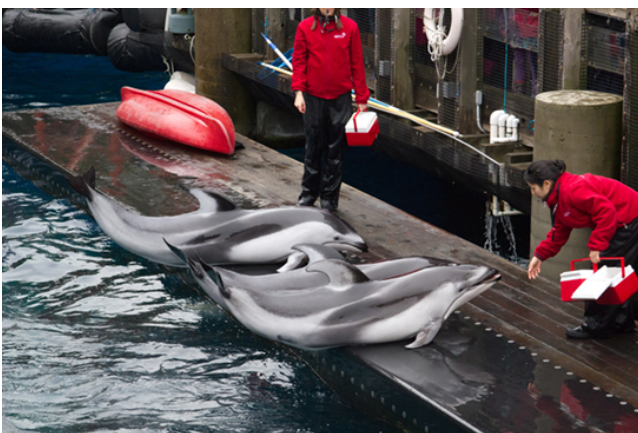
In the classic Pavlovian response, the dog becomes conditioned to associate the ringing of the bell with food.

It had been thought by some scientists that this type of conditioning required multiple exposures to the paired stimulus and response, but it is now known that this is not necessary in all cases, and that some

conditioning can be learned in a single pairing experiment. Classical conditioning is a major tenet of behaviorism, a branch of psychological philosophy that proposes that all actions, thoughts, and emotions of living things are behaviors that can be treated by behavior modification and changes in the environment.

## Operant Conditioning

In **operant conditioning**, the conditioned behavior is gradually modified by its consequences as the animal responds to the stimulus. A major proponent of such conditioning was psychologist B.F. Skinner, the inventor of the Skinner box. Skinner put rats in his boxes that contained a lever that would dispense food to the rat when depressed. While initially the rat would push the lever a few times by accident, it eventually associated pushing the lever with getting the food. This type of learning is an example of operant conditioning. Operant learning is the basis of most animal training. The conditioned behavior is continually modified by positive or negative reinforcement, often a reward such as food or some type of punishment, respectively. In this way, the animal is conditioned to associate a type of behavior with the punishment or reward, and, over time, can be induced to perform behaviors that they would not have done in the wild, such as the “tricks” dolphins perform at marine amusement park shows ([\[link\]](#)).



The training of dolphins by

rewarding them with food is an example of positive reinforcement operant conditioning. (credit: Roland Tanglao)

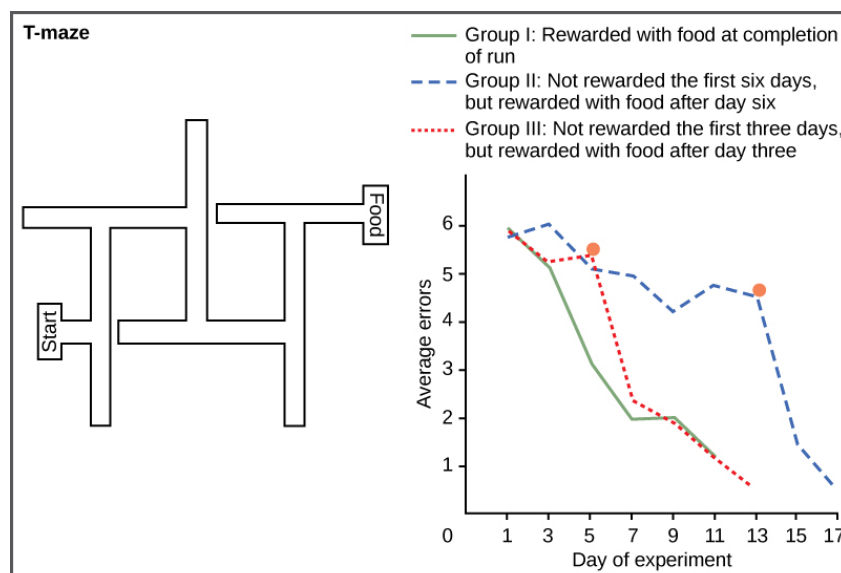
## Cognitive Learning

Classical and operant conditioning are inefficient ways for humans and other intelligent animals to learn. Some primates, including humans, are able to learn by imitating the behavior of others and by taking instructions. The development of complex language by humans has made **cognitive learning**, the manipulation of information using the mind, the most prominent method of human learning. In fact, that is how students are learning right now by reading this book. As students read, they can make mental images of objects or organisms and imagine changes to them, or behaviors by them, and anticipate the consequences. In addition to visual processing, cognitive learning is also enhanced by remembering past experiences, touching physical objects, hearing sounds, tasting food, and a variety of other sensory-based inputs. Cognitive learning is so powerful that it can be used to understand conditioning in detail. In the reverse scenario, conditioning cannot help someone learn about cognition.

Classic work on cognitive learning was done by Wolfgang Köhler with chimpanzees. He demonstrated that these animals were capable of abstract thought by showing that they could learn how to solve a puzzle. When a banana was hung in their cage too high for them to reach, and several boxes were placed randomly on the floor, some of the chimps were able to stack the boxes one on top of the other, climb on top of them, and get the banana. This implies that they could visualize the result of stacking the boxes even before they had performed the action. This type of learning is much more powerful and versatile than conditioning.

Cognitive learning is not limited to primates, although they are the most efficient in using it. Maze running experiments done with rats by H.C. Blodgett in the 1920s were the first to show cognitive skills in a simple

mammal. The motivation for the animals to work their way through the maze was a piece of food at its end. In these studies, the animals in Group I were run in one trial per day and had food available to them each day on completion of the run ([link](#)). Group II rats were not fed in the maze for the first six days and then subsequent runs were done with food for several days after. Group III rats had food available on the third day and every day thereafter. The results were that the control rats, Group I, learned quickly, and figured out how to run the maze in seven days. Group III did not learn much during the three days without food, but rapidly caught up to the control group when given the food reward. Group II learned very slowly for the six days with no reward to motivate them, and they did not begin to catch up to the control group until the day food was given, and then it took two days longer to learn the maze.



Redrawn after H. C. Blodgett, The effect of the introduction of reward upon the maze performance of rats. Univ. Calif. Publ. Psychol., 1929, 4, No. 8, pages 117 and 120.

Group I (the green solid line) found food at the end of each trial, group II (the blue dashed line) did not find food for the first 6 days, and group III (the red dotted line) did not find food during runs on the first three days. Notice that rats given food earlier learned faster and eventually caught up to the control group. The

orange dots on the group II and III lines show the days when food rewards were added to the mazes.

It may not be immediately obvious that this type of learning is different than conditioning. Although one might be tempted to believe that the rats simply learned how to find their way through a conditioned series of right and left turns, E.C. Tolman proved a decade later that the rats were making a representation of the maze in their minds, which he called a “cognitive map.” This was an early demonstration of the power of cognitive learning and how these abilities were not just limited to humans.

## **Sociobiology**

Sociobiology is an interdisciplinary science originally popularized by social insect researcher E.O. Wilson in the 1970s. Wilson defined the science as “the extension of population biology and evolutionary theory to social organization.”<sup>[footnote]</sup> The main thrust of sociobiology is that animal and human behavior, including aggressiveness and other social interactions, can be explained almost solely in terms of genetics and natural selection. This science is controversial; noted scientist such as the late Stephen Jay Gould criticized the approach for ignoring the environmental effects on behavior. This is another example of the “nature versus nurture” debate of the role of genetics versus the role of environment in determining an organism’s characteristics.

Edward O. Wilson. *On Human Nature* (1978; repr., Cambridge: Harvard University Press, 2004), xx.

Sociobiology also links genes with behaviors and has been associated with “biological determinism,” the belief that all behaviors are hardwired into our genes. No one disputes that certain behaviors can be inherited and that natural selection plays a role retaining them. It is the application of such principles to human behavior that sparks this controversy, which remains active today.

## Section Summary

Behaviors are responses to stimuli. They can either be instinctual/innate behaviors, which are not influenced by the environment, or learned behaviors, which are influenced by environmental changes. Instinctual behaviors include mating systems and methods of communication. Learned behaviors include imprinting and habituation, conditioning, and, most powerfully, cognitive learning. Although the connection between behavior, genetics, and evolution is well established, the explanation of human behavior as entirely genetic is controversial.

## Review Questions

### Exercise:

#### Problem:

The ability of rats to learn how to run a maze is an example of \_\_\_\_\_.

- a. imprinting
- b. classical conditioning
- c. operant conditioning
- d. cognitive learning

---

#### Solution:

D

### Exercise:

**Problem:** The training of animals usually involves \_\_\_\_\_.

- a. imprinting
- b. classical conditioning
- c. operant conditioning
- d. cognitive learning

---

**Solution:**

C

**Exercise:****Problem:**

The sacrifice of the life of an individual so that the genes of relatives may be passed on is called \_\_\_\_\_.

- a. operant learning
- b. kin selection
- c. kinesis
- d. imprinting

---

**Solution:**

B

**Free Response****Exercise:****Problem:**

Describe Pavlov's dog experiments as an example of classical conditioning.

---

**Solution:**

Dogs salivated in response to food. This was the unconditioned stimulus and response. Dogs exposed to food had a bell rung repeatedly at the same time, eventually learning to associate the bell with food. Over time, the dogs would salivate when the bell was rung, even in the absence of food. Thus, the bell became the conditioned



stimulus, and the salivation in response to the bell became the conditioned response.

## **Glossary**

aggressive display

visual display by a species member to discourage other members of the same species or different species

behavior

change in an organism's activities in response to a stimulus

behavioral biology

study of the biology and evolution of behavior

classical conditioning

association of a specific stimulus and response through conditioning

cognitive learning

knowledge and skills acquired by the manipulation of information in the mind

conditioned behavior

behavior that becomes associated with a specific stimulus through conditioning

courtship display

visual display used to attract a mate

distraction display

visual display used to distract predators away from a nesting site

ethology

biological study of animal behavior

fixed action pattern

series of instinctual behaviors that, once initiated, always goes to completion regardless of changes in the environment

foraging

behaviors species use to find food

habituation

ability of a species to ignore repeated stimuli that have no consequence

imprinting

identification of parents by newborns as the first organism they see after birth

innate behavior

instinctual behavior that is not altered by changes in the environment

intersexual selection

selection of a desirable mate of the opposite sex

intrasexual selection

competition between members of the same sex for a mate

kin selection

sacrificing one's own life so that one's genes will be passed on to future generations by relatives

kinesis

undirected movement of an organism in response to a stimulus

learned behavior

behavior that responds to changes in the environment

migration

long-range seasonal movement of animal species

monogamy

mating system whereby one male and one female remain coupled for at least one mating season

operant conditioning

learned behaviors in response to positive and/or negative reinforcement

polyandry

mating system where one female mates with many males

polygyny

mating system where one male mates with many females

reflex action

action in response to direct physical stimulation of a nerve

signal

method of communication between animals including those obtained by the senses of smell, hearing, sight, or touch

taxis

directed movement in response to a stimulus

## Introduction

class="introduction"

In the southwestern United States, rainy weather causes an increase in production of pinyon nuts, causing the deer mouse population to explode.

Deer mice may carry a virus called *Sin Nombre*

(a hantavirus) that causes respiratory disease in humans and has a high fatality rate.

In 1992–1993, wet *El*

*Niño* weather caused a *Sin Nombre* epidemic. Navajo

healers, who  
were aware  
of the link  
between this  
disease and  
weather,  
predicted the  
outbreak.  
(credit  
"highway":  
modification  
of work by  
Phillip  
Capper;  
credit  
"mouse":  
modification  
of work by  
USFWS)



In 1993, an interesting example of ecosystem dynamics occurred when a rare lung disease struck inhabitants of the southwestern United States. This disease had an alarming rate of fatalities, killing more than half of early patients, many of whom were Native Americans. These formerly healthy

young adults died from complete respiratory failure. The disease was unknown, and the Centers for Disease Control (CDC), the United States government agency responsible for managing potential epidemics, was brought in to investigate. The scientists could have learned about the disease had they known to talk with the Navajo healers who lived in the area and who had observed the connection between rainfall and mice populations, thereby predicting the 1993 outbreak.

The cause of the disease, determined within a few weeks by the CDC investigators, was the hantavirus known as *Sin Nombre*, the virus with “no name.” With insights from traditional Navajo medicine, scientists were able to characterize the disease rapidly and institute effective health measures to prevent its spread. This example illustrates the importance of understanding the complexities of ecosystems and how they respond to changes in the environment.

## Ecology of Ecosystems

By the end of this section, you will be able to:

- Describe the basic types of ecosystems on Earth
- Explain the methods that ecologists use to study ecosystem structure and dynamics
- Identify the different methods of ecosystem modeling
- Differentiate between food chains and food webs and recognize the importance of each

Life in an ecosystem is often about competition for limited resources, a characteristic of the theory of natural selection. Competition in communities (all living things within specific habitats) is observed both within species and among different species. The resources for which organisms compete include organic material from living or previously living organisms, sunlight, and mineral nutrients, which provide the energy for living processes and the matter to make up organisms' physical structures. Other critical factors influencing community dynamics are the components of its physical and geographic environment: a habitat's latitude, amount of rainfall, topography (elevation), and available species. These are all important environmental variables that determine which organisms can exist within a particular area.

An **ecosystem** is a community of living organisms and their interactions with their abiotic (non-living) environment. Ecosystems can be small, such as the tide pools found near the rocky shores of many oceans, or large, such as the Amazon Rainforest in Brazil ([\[link\]](#)).



(a)



(b)

A (a) tidal pool ecosystem in Matinicus Island in Maine is a small ecosystem, while the (b) Amazon Rainforest in Brazil is a large ecosystem. (credit a: modification of work by “takomabibelot”/Flickr; credit b: modification of work by Ivan Mlinaric)

There are three broad categories of ecosystems based on their general environment: freshwater, ocean water, and terrestrial. Within these broad categories are individual ecosystem types based on the organisms present and the type of environmental habitat.

Ocean ecosystems are the most common, comprising 75 percent of the Earth's surface and consisting of three basic types: shallow ocean, deep ocean water, and deep ocean surfaces (the low depth areas of the deep oceans). The shallow ocean ecosystems include extremely biodiverse coral reef ecosystems, and the deep ocean surface is known for its large numbers of plankton and krill (small crustaceans) that support it. These two environments are especially important to aerobic respirators worldwide as the phytoplankton perform 40 percent of all photosynthesis on Earth. Although not as diverse as the other two, deep ocean ecosystems contain a wide variety of marine organisms. Such ecosystems exist even at the bottom of the ocean where light is unable to penetrate through the water.

Freshwater ecosystems are the rarest, occurring on only 1.8 percent of the Earth's surface. Lakes, rivers, streams, and springs comprise these systems; they are quite diverse, and they support a variety of fish, amphibians, reptiles, insects, phytoplankton, fungi, and bacteria.

Terrestrial ecosystems, also known for their diversity, are grouped into large categories called biomes, such as tropical rain forests, savannas, deserts, coniferous forests, deciduous forests, and tundra. Grouping these ecosystems into just a few biome categories obscures the great diversity of the individual ecosystems within them. For example, there is great variation in desert vegetation: the saguaro cacti and other plant life in the Sonoran Desert, in the United States, are relatively abundant compared to the



desolate rocky desert of Boa Vista, an island off the coast of Western Africa ([link](#)).



(a)



(b)

Desert ecosystems, like all ecosystems, can vary greatly. The desert in (a) Saguaro National Park, Arizona, has abundant plant life, while the rocky desert of (b) Boa Vista island, Cape Verde, Africa, is devoid of plant life. (credit a: modification of work by Jay Galvin; credit b: modification of work by Ingo Wölbern)

Ecosystems are complex with many interacting parts. They are routinely exposed to various disturbances, or changes in the environment that effect their compositions: yearly variations in rainfall and temperature and the slower processes of plant growth, which may take several years. Many of these disturbances are a result of natural processes. For example, when lightning causes a forest fire and destroys part of a forest ecosystem, the ground is eventually populated by grasses, then by bushes and shrubs, and later by mature trees, restoring the forest to its former state. The impact of environmental disturbances caused by human activities is as important as the changes wrought by natural processes. Human agricultural practices, air pollution, acid rain, global deforestation, overfishing, eutrophication, oil spills, and illegal dumping on land and into the ocean are all issues of concern to conservationists.

**Equilibrium** is the steady state of an ecosystem where all organisms are in balance with their environment and with each other. In ecology, two

parameters are used to measure changes in ecosystems: resistance and resilience. The ability of an ecosystem to remain at equilibrium in spite of disturbances is called **resistance**. The speed at which an ecosystem recovers equilibrium after being disturbed, called its **resilience**. Ecosystem resistance and resilience are especially important when considering human impact. The nature of an ecosystem may change to such a degree that it can lose its resilience entirely. This process can lead to the complete destruction or irreversible altering of the ecosystem.

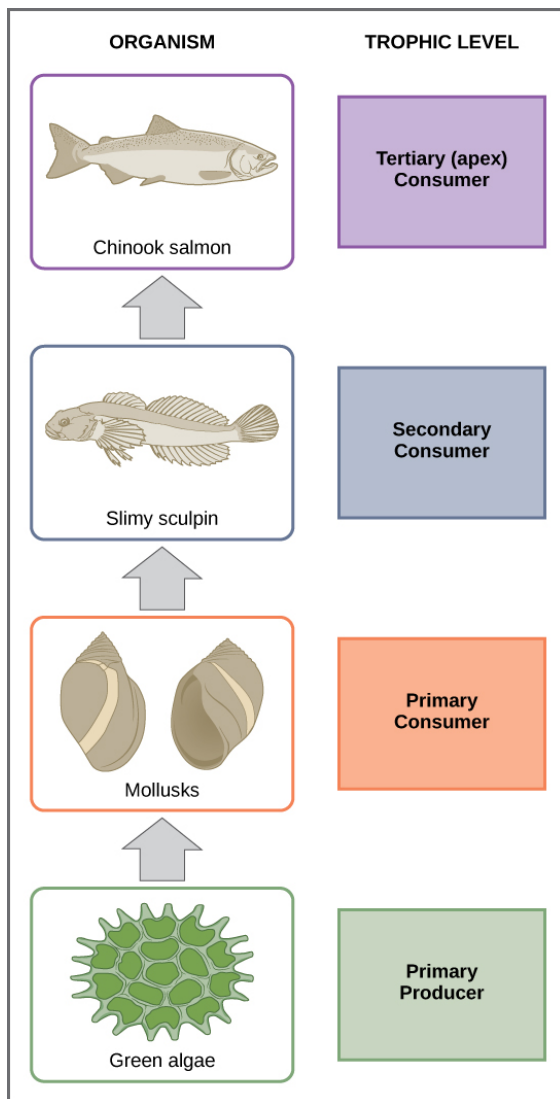
## Food Chains and Food Webs

The term “food chain” is sometimes used metaphorically to describe human social situations. In this sense, food chains are thought of as a competition for survival, such as “who eats whom?” Someone eats and someone is eaten. Therefore, it is not surprising that in our competitive “dog-eat-dog” society, individuals who are considered successful are seen as being at the top of the food chain, consuming all others for their benefit, whereas the less successful are seen as being at the bottom.

The scientific understanding of a food chain is more precise than in its everyday usage. In ecology, a **food chain** is a linear sequence of organisms through which nutrients and energy pass: primary producers, primary consumers, and higher-level consumers are used to describe ecosystem structure and dynamics. There is a single path through the chain. Each organism in a food chain occupies what is called a **trophic level**. Depending on their role as producers or consumers, species or groups of species can be assigned to various trophic levels.

In many ecosystems, the bottom of the food chain consists of photosynthetic organisms (plants and/or phytoplankton), which are called **primary producers**. The organisms that consume the primary producers are herbivores: the **primary consumers**. **Secondary consumers** are usually carnivores that eat the primary consumers. **Tertiary consumers** are carnivores that eat other carnivores. Higher-level consumers feed on the next lower trophic levels, and so on, up to the organisms at the top of the food chain: the **apex consumers**. In the Lake Ontario food chain shown in

[[link](#)], the Chinook salmon is the apex consumer at the top of this food chain.

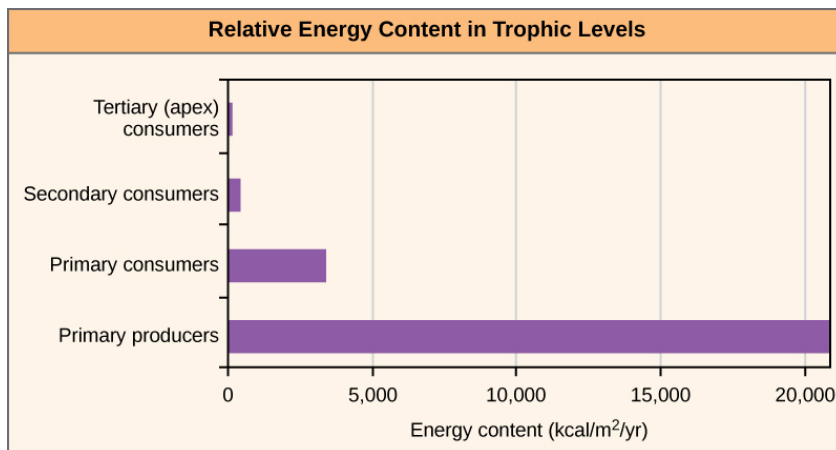


These are the trophic levels of a food chain in Lake Ontario at the United States-Canada border. Energy and nutrients flow from photosynthetic green algae at the bottom to the top of the

## food chain: the Chinook salmon.

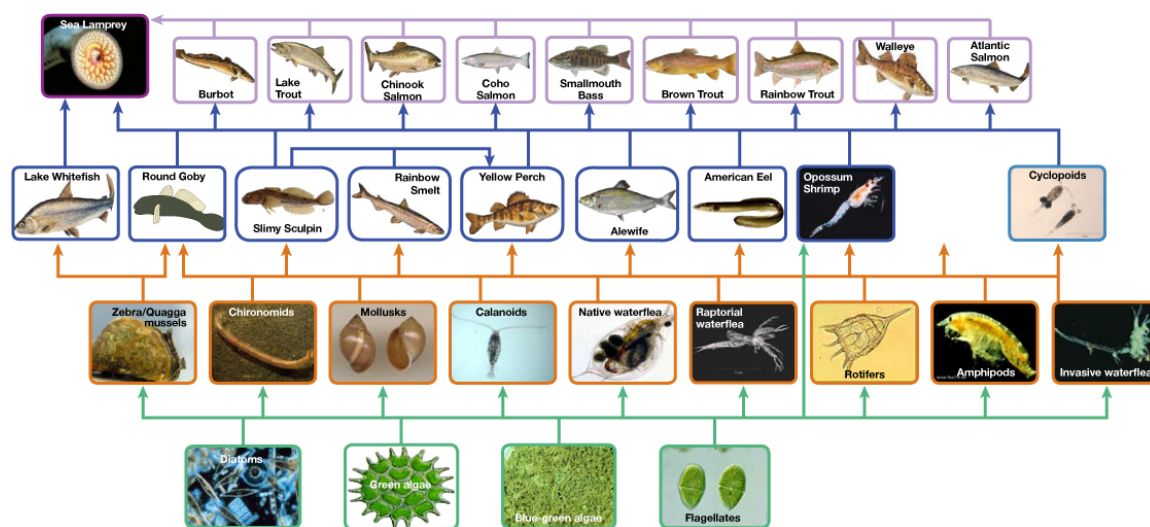
One major factor that limits the length of food chains is energy. Energy is lost as heat between each trophic level due to the second law of thermodynamics. Thus, after a limited number of trophic energy transfers, the amount of energy remaining in the food chain may not be great enough to support viable populations at yet a higher trophic level.

The loss of energy between trophic levels is illustrated by the pioneering studies of Howard T. Odum in the Silver Springs, Florida, ecosystem in the 1940s ([link](#)). The primary producers generated 20,819 kcal/m<sup>2</sup>/yr (kilocalories per square meter per year), the primary consumers generated 3368 kcal/m<sup>2</sup>/yr, the secondary consumers generated 383 kcal/m<sup>2</sup>/yr, and the tertiary consumers only generated 21 kcal/m<sup>2</sup>/yr. Thus, there is little energy remaining for another level of consumers in this ecosystem.



The relative energy in trophic levels in a Silver Springs, Florida, ecosystem is shown. Each trophic level has less energy available and supports fewer organisms at the next level.

There is a one problem when using food chains to accurately describe most ecosystems. Even when all organisms are grouped into appropriate trophic levels, some of these organisms can feed on species from more than one trophic level; likewise, some of these organisms can be eaten by species from multiple trophic levels. In other words, the linear model of ecosystems, the food chain, is not completely descriptive of ecosystem structure. A holistic model—which accounts for all the interactions between different species and their complex interconnected relationships with each other and with the environment—is a more accurate and descriptive model for ecosystems. A **food web** is a graphic representation of a holistic, non-linear web of primary producers, primary consumers, and higher-level consumers used to describe ecosystem structure and dynamics ([\[link\]](#)).



This food web shows the interactions between organisms across trophic levels in the Lake Ontario ecosystem. Primary producers are outlined in green, primary consumers in orange, secondary consumers in blue, and tertiary (apex) consumers in purple. Arrows point from an organism that is consumed to the organism that consumes it. Notice how some lines point to more than one trophic level. For example, the opossum shrimp eats both primary producers and primary consumers. (credit: NOAA, GLERL)

A comparison of the two types of structural ecosystem models shows strength in both. Food chains are more flexible for analytical modeling, are easier to follow, and are easier to experiment with, whereas food web models more accurately represent ecosystem structure and dynamics, and data can be directly used as input for simulation modeling.

**Note:**

Link to Learning



Head to this [online interactive simulator](#) to investigate food web function. In the *Interactive Labs* box, under Food Web, click **Step 1**. Read the instructions first, and then click **Step 2** for additional instructions. When you are ready to create a simulation, in the upper-right corner of the *Interactive Labs* box, click **OPEN SIMULATOR**.

Two general types of food webs are often shown interacting within a single ecosystem. A **grazing food web** (such as the Lake Ontario food web in [\[link\]](#)) has plants or other photosynthetic organisms at its base, followed by herbivores and various carnivores. A **detrital food web** consists of a base of organisms that feed on decaying organic matter (dead organisms), called decomposers or detritivores. These organisms are usually bacteria or fungi that recycle organic material back into the biotic part of the ecosystem as they themselves are consumed by other organisms. As all ecosystems require a method to recycle material from dead organisms, most grazing food webs have an associated detrital food web. For example, in a meadow ecosystem, plants may support a grazing food web of different organisms, primary and other levels of consumers, while at the same time supporting a

detrital food web of bacteria, fungi, and detritivorous invertebrates feeding off dead plants and animals.

**Note:**

**Evolution Connection**

**Three-spined Stickleback**

It is well established by the theory of natural selection that changes in the environment play a major role in the evolution of species within an ecosystem. However, little is known about how the evolution of species within an ecosystem can alter the ecosystem environment. In 2009, Dr. Luke Harmon, from the University of Idaho in Moscow, published a paper that for the first time showed that the evolution of organisms into subspecies can have direct effects on their ecosystem environment. [\[footnote\]](#)  
*Nature* (Vol. 458, April 1, 2009)

The three-spines stickleback (*Gasterosteus aculeatus*) is a freshwater fish that evolved from a saltwater fish to live in freshwater lakes about 10,000 years ago, which is considered a recent development in evolutionary time ([\[link\]](#)). Over the last 10,000 years, these freshwater fish then became isolated from each other in different lakes. Depending on which lake population was studied, findings showed that these sticklebacks then either remained as one species or evolved into two species. The divergence of species was made possible by their use of different areas of the pond for feeding called micro niches.

Dr. Harmon and his team created artificial pond microcosms in 250-gallon tanks and added muck from freshwater ponds as a source of zooplankton and other invertebrates to sustain the fish. In different experimental tanks they introduced one species of stickleback from either a single-species or double-species lake.

Over time, the team observed that some of the tanks bloomed with algae while others did not. This puzzled the scientists, and they decided to measure the water's dissolved organic carbon (DOC), which consists of mostly large molecules of decaying organic matter that give pond-water its slightly brownish color. It turned out that the water from the tanks with two-species fish contained larger particles of DOC (and hence darker water) than water with single-species fish. This increase in DOC blocked



the sunlight and prevented algal blooming. Conversely, the water from the single-species tank contained smaller DOC particles, allowing more sunlight penetration to fuel the algal blooms.

This change in the environment, which is due to the different feeding habits of the stickleback species in each lake type, probably has a great impact on the survival of other species in these ecosystems, especially other photosynthetic organisms. Thus, the study shows that, at least in these ecosystems, the environment and the evolution of populations have reciprocal effects that may now be factored into simulation models.



The three-spined stickleback evolved from a saltwater fish to freshwater fish. (credit: Barrett Paul, USFWS)

## Research into Ecosystem Dynamics: Ecosystem Experimentation and Modeling

The study of the changes in ecosystem structure caused by changes in the environment (disturbances) or by internal forces is called **ecosystem dynamics**. Ecosystems are characterized using a variety of research methodologies. Some ecologists study ecosystems using controlled experimental systems, while some study entire ecosystems in their natural state, and others use both approaches.



A **holistic ecosystem model** attempts to quantify the composition, interaction, and dynamics of entire ecosystems; it is the most representative of the ecosystem in its natural state. A food web is an example of a holistic ecosystem model. However, this type of study is limited by time and expense, as well as the fact that it is neither feasible nor ethical to do experiments on large natural ecosystems. To quantify all different species in an ecosystem and the dynamics in their habitat is difficult, especially when studying large habitats such as the Amazon Rainforest, which covers 1.4 billion acres (5.5 million km<sup>2</sup>) of the Earth's surface.

For these reasons, scientists study ecosystems under more controlled conditions. Experimental systems usually involve either partitioning a part of a natural ecosystem that can be used for experiments, termed a **mesocosm**, or by re-creating an ecosystem entirely in an indoor or outdoor laboratory environment, which is referred to as a **microcosm**. A major limitation to these approaches is that removing individual organisms from their natural ecosystem or altering a natural ecosystem through partitioning may change the dynamics of the ecosystem. These changes are often due to differences in species numbers and diversity and also to environment alterations caused by partitioning (mesocosm) or re-creating (microcosm) the natural habitat. Thus, these types of experiments are not totally predictive of changes that would occur in the ecosystem from which they were gathered.

As both of these approaches have their limitations, some ecologists suggest that results from these experimental systems should be used only in conjunction with holistic ecosystem studies to obtain the most representative data about ecosystem structure, function, and dynamics.

Scientists use the data generated by these experimental studies to develop ecosystem models that demonstrate the structure and dynamics of ecosystems. Three basic types of ecosystem modeling are routinely used in research and ecosystem management: a conceptual model, an analytical model, and a simulation model. A **conceptual model** is an ecosystem model that consists of flow charts to show interactions of different compartments of the living and nonliving components of the ecosystem. A conceptual model describes ecosystem structure and dynamics and shows

how environmental disturbances affect the ecosystem; however, its ability to predict the effects of these disturbances is limited. Analytical and simulation models, in contrast, are mathematical methods of describing ecosystems that are indeed capable of predicting the effects of potential environmental changes without direct experimentation, although with some limitations as to accuracy. An **analytical model** is an ecosystem model that is created using simple mathematical formulas to predict the effects of environmental disturbances on ecosystem structure and dynamics. A **simulation model** is an ecosystem model that is created using complex computer algorithms to holistically model ecosystems and to predict the effects of environmental disturbances on ecosystem structure and dynamics. Ideally, these models are accurate enough to determine which components of the ecosystem are particularly sensitive to disturbances, and they can serve as a guide to ecosystem managers (such as conservation ecologists or fisheries biologists) in the practical maintenance of ecosystem health.

## Conceptual Models

Conceptual models are useful for describing ecosystem structure and dynamics and for demonstrating the relationships between different organisms in a community and their environment. Conceptual models are usually depicted graphically as flow charts. The organisms and their resources are grouped into specific compartments with arrows showing the relationship and transfer of energy or nutrients between them. Thus, these diagrams are sometimes called compartment models.

To model the cycling of mineral nutrients, organic and inorganic nutrients are subdivided into those that are bioavailable (ready to be incorporated into biological macromolecules) and those that are not. For example, in a terrestrial ecosystem near a deposit of coal, carbon will be available to the plants of this ecosystem as carbon dioxide gas in a short-term period, not from the carbon-rich coal itself. However, over a longer period, microorganisms capable of digesting coal will incorporate its carbon or release it as natural gas (methane,  $\text{CH}_4$ ), changing this unavailable organic source into an available one. This conversion is greatly accelerated by the combustion of fossil fuels by humans, which releases large amounts of

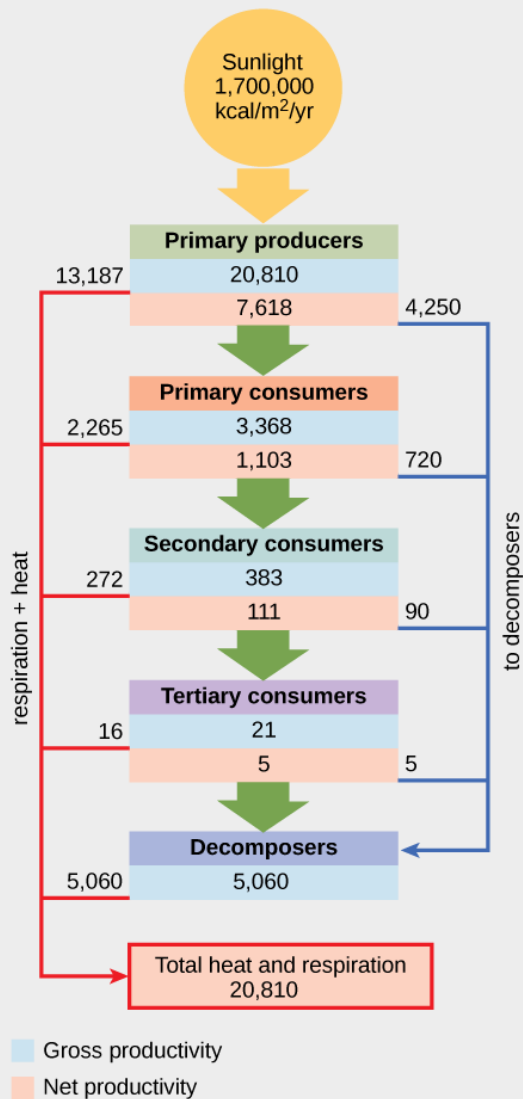
carbon dioxide into the atmosphere. This is thought to be a major factor in the rise of the atmospheric carbon dioxide levels in the industrial age. The carbon dioxide released from burning fossil fuels is produced faster than photosynthetic organisms can use it. This process is intensified by the reduction of photosynthetic trees because of worldwide deforestation. Most scientists agree that high atmospheric carbon dioxide is a major cause of global climate change.

Conceptual models are also used to show the flow of energy through particular ecosystems. [\[link\]](#) is based on Howard T. Odum's classical study of the Silver Springs, Florida, holistic ecosystem in the mid-twentieth century.<sup>[\[footnote\]](#)</sup> This study shows the energy content and transfer between various ecosystem compartments.

Howard T. Odum, "Trophic Structure and Productivity of Silver Springs, Florida," *Ecological Monographs* 27, no. 1 (1957): 47–112.

**Note:**

Art Connection



This conceptual model shows the flow of energy through a spring ecosystem in Silver Springs, Florida. Notice that the energy decreases with each increase in trophic level.

Why do you think the value for gross productivity of the primary producers is the same as the value for total heat and respiration (20,810 kcal/m<sup>2</sup>/yr)?

## Analytical and Simulation Models

The major limitation of conceptual models is their inability to predict the consequences of changes in ecosystem species and/or environment. Ecosystems are dynamic entities and subject to a variety of abiotic and biotic disturbances caused by natural forces and/or human activity. Ecosystems altered from their initial equilibrium state can often recover from such disturbances and return to a state of equilibrium. As most ecosystems are subject to periodic disturbances and are often in a state of change, they are usually either moving toward or away from their equilibrium state. There are many of these equilibrium states among the various components of an ecosystem, which affects the ecosystem overall. Furthermore, as humans have the ability to greatly and rapidly alter the species content and habitat of an ecosystem, the need for predictive models that enable understanding of how ecosystems respond to these changes becomes more crucial.

Analytical models often use simple, linear components of ecosystems, such as food chains, and are known to be complex mathematically; therefore, they require a significant amount of mathematical knowledge and expertise. Although analytical models have great potential, their simplification of complex ecosystems is thought to limit their accuracy. Simulation models that use computer programs are better able to deal with the complexities of ecosystem structure.

A recent development in simulation modeling uses supercomputers to create and run individual-based simulations, which accounts for the behavior of individual organisms and their effects on the ecosystem as a whole. These simulations are considered to be the most accurate and predictive of the complex responses of ecosystems to disturbances.

### **Note:**

[Link to Learning](#)



Visit [The Darwin Project](#) to view a variety of ecosystem models.

## Section Summary

Ecosystems exist on land, at sea, in the air, and underground. Different ways of modeling ecosystems are necessary to understand how environmental disturbances will affect ecosystem structure and dynamics. Conceptual models are useful to show the general relationships between organisms and the flow of materials or energy between them. Analytical models are used to describe linear food chains, and simulation models work best with holistic food webs.

## Art Connections

### Exercise:

#### Problem:

[\[link\]](#) Why do you think the value for gross productivity of the primary producers is the same as the value for total heat and respiration (20,810 kcal/m<sup>2</sup>/yr)?

---

#### Solution:

[\[link\]](#) According to the first law of thermodynamics, energy can neither be created nor destroyed. Eventually, all energy consumed by living systems is lost as heat or used for respiration, and the total energy output of the system must equal the energy that went into it.

## Review Questions

### Exercise:

#### Problem:

The ability of an ecosystem to return to its equilibrium state after an environmental disturbance is called \_\_\_\_\_.

- a. resistance
- b. restoration
- c. reformation
- d. resilience

---

#### Solution:

D

### Exercise:

#### Problem:

A re-created ecosystem in a laboratory environment is known as a \_\_\_\_\_.

- a. mesocosm
- b. simulation
- c. microcosm
- d. reproduction

---

#### Solution:

C

### Exercise:

**Problem:**Decomposers are associated with which class of food web?

- a. grazing
- b. detrital
- c. inverted
- d. aquatic

---

**Solution:**

B

**Exercise:**

**Problem:**

The primary producers in an ocean grazing food web are usually \_\_\_\_\_.

- a. plants
- b. animals
- c. fungi
- d. phytoplankton

---

**Solution:**

D

**Exercise:**

**Problem:**

What term describes the use of mathematical equations in the modeling of linear aspects of ecosystems?

- a. analytical modeling
- b. simulation modeling
- c. conceptual modeling
- d. individual-based modeling

---

**Solution:**



A

**Exercise:**

**Problem:**

The position of an organism along a food chain is known as its \_\_\_\_\_.

- a. locus
- b. location
- c. trophic level
- d. microcosm

---

**Solution:**

C

**Free Response**

**Exercise:**

**Problem:**

Compare and contrast food chains and food webs. What are the strengths of each concept in describing ecosystems?

---

**Solution:**

Food webs show interacting groups of different species and their many interconnections with each other and the environment. Food chains are linear aspects of food webs that describe the succession of organisms consuming one another at defined trophic levels. Food webs are a more accurate representation of the structure and dynamics of an ecosystem. Food chains are easier to model and use for experimental studies.

**Exercise:**

**Problem:** Describe freshwater, ocean, and terrestrial ecosystems.

---

**Solution:**

Freshwater ecosystems are the rarest, but have great diversity of freshwater fish and other aquatic life. Ocean ecosystems are the most common and are responsible for much of the photosynthesis that occurs on Earth. Terrestrial ecosystems are very diverse; they are grouped based on their species and environment (biome), which includes forests, deserts, and tundras.

**Exercise:**

**Problem:**

Compare grazing and detrital food webs. Why would they both be present in the same ecosystem?

---

**Solution:**

Grazing food webs have a primary producer at their base, which is either a plant for terrestrial ecosystems or a phytoplankton for aquatic ecosystems. The producers pass their energy to the various trophic levels of consumers. At the base of detrital food webs are the decomposers, which pass this energy to a variety of other consumers. Detrital food webs are important for the health of many grazing food webs because they eliminate dead and decaying organic material, thus, clearing space for new organisms and removing potential causes of disease. By breaking down dead organic matter, decomposers also make mineral nutrients available to primary producers; this process is a vital link in nutrient cycling.

## Glossary

analytical model

ecosystem model that is created with mathematical formulas to predict the effects of environmental disturbances on ecosystem structure and

dynamics

apex consumer

organism at the top of the food chain

conceptual model

(also, compartment models) ecosystem model that consists of flow charts that show the interactions of different compartments of the living and non-living components of the ecosystem

detrital food web

type of food web in which the primary consumers consist of decomposers; these are often associated with grazing food webs within the same ecosystem

ecosystem

community of living organisms and their interactions with their abiotic environment

ecosystem dynamics

study of the changes in ecosystem structure caused by changes in the environment or internal forces

equilibrium

steady state of an ecosystem where all organisms are in balance with their environment and each other

food chain

linear representation of a chain of primary producers, primary consumers, and higher-level consumers used to describe ecosystem structure and dynamics

food web

graphic representation of a holistic, non-linear web of primary producers, primary consumers, and higher-level consumers used to describe ecosystem structure and dynamics

grazing food web

type of food web in which the primary producers are either plants on land or phytoplankton in the water; often associated with a detrital food web within the same ecosystem

holistic ecosystem model

study that attempts to quantify the composition, interactions, and dynamics of entire ecosystems; often limited by economic and logistical difficulties, depending on the ecosystem

mesocosm

portion of a natural ecosystem to be used for experiments

microcosm

re-creation of natural ecosystems entirely in a laboratory environment to be used for experiments

primary consumer

trophic level that obtains its energy from the primary producers of an ecosystem

primary producer

trophic level that obtains its energy from sunlight, inorganic chemicals, or dead and/or decaying organic material

resilience (ecological)

speed at which an ecosystem recovers equilibrium after being disturbed

resistance (ecological)

ability of an ecosystem to remain at equilibrium in spite of disturbances

secondary consumer

usually a carnivore that eat primary consumers

simulation model

ecosystem model that is created with computer programs to holistically model ecosystems and to predict the effects of environmental

disturbances on ecosystem structure and dynamics

tertiary consumer

carnivore that eat other carnivores

trophic level

position of a species or group of species in a food chain or a food web

## Energy Flow through Ecosystems

By the end of this section, you will be able to:

- Describe how organisms acquire energy in a food web and in associated food chains
- Explain how the efficiency of energy transfers between trophic levels affects ecosystem structure and dynamics
- Discuss trophic levels and how ecological pyramids are used to model them

All living things require energy in one form or another. Energy is required by most complex metabolic pathways (often in the form of adenosine triphosphate, ATP), especially those responsible for building large molecules from smaller compounds, and life itself is an energy-driven process. Living organisms would not be able to assemble macromolecules (proteins, lipids, nucleic acids, and complex carbohydrates) from their monomeric subunits without a constant energy input.

It is important to understand how organisms acquire energy and how that energy is passed from one organism to another through food webs and their constituent food chains. Food webs illustrate how energy flows directionally through ecosystems, including how efficiently organisms acquire it, use it, and how much remains for use by other organisms of the food web.

## How Organisms Acquire Energy in a Food Web

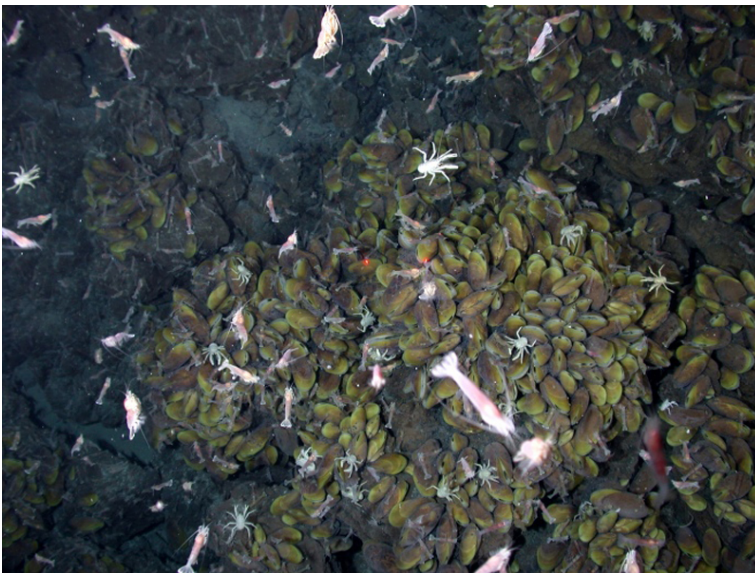
Energy is acquired by living things in three ways: photosynthesis, chemosynthesis, and the consumption and digestion of other living or previously living organisms by heterotrophs.

Photosynthetic and chemosynthetic organisms are both grouped into a category known as autotrophs: organisms capable of synthesizing their own food (more specifically, capable of using inorganic carbon as a carbon source). Photosynthetic autotrophs (photoautotrophs) use sunlight as an energy source, whereas chemosynthetic autotrophs (chemoautotrophs) use inorganic molecules as an energy source. Autotrophs are critical for all

ecosystems. Without these organisms, energy would not be available to other living organisms and life itself would not be possible.

Photoautotrophs, such as plants, algae, and photosynthetic bacteria, serve as the energy source for a majority of the world's ecosystems. These ecosystems are often described by grazing food webs. Photoautotrophs harness the solar energy of the sun by converting it to chemical energy in the form of ATP (and NADP). The energy stored in ATP is used to synthesize complex organic molecules, such as glucose.

**Chemoautotrophs** are primarily bacteria that are found in rare ecosystems where sunlight is not available, such as in those associated with dark caves or hydrothermal vents at the bottom of the ocean ([\[link\]](#)). Many chemoautotrophs in hydrothermal vents use hydrogen sulfide ( $H_2S$ ), which is released from the vents as a source of chemical energy. This allows chemoautotrophs to synthesize complex organic molecules, such as glucose, for their own energy and in turn supplies energy to the rest of the ecosystem.



Swimming shrimp, a few squat lobsters, and hundreds of vent mussels are seen at a hydrothermal vent at the bottom of the

ocean. As no sunlight penetrates to this depth, the ecosystem is supported by chemoautotrophic bacteria and organic material that sinks from the ocean's surface. This picture was taken in 2006 at the submerged NW Eifuku volcano off the coast of Japan by the National Oceanic and Atmospheric Administration (NOAA). The summit of this highly active volcano lies 1535 m below the surface.

## Productivity within Trophic Levels

Productivity within an ecosystem can be defined as the percentage of energy entering the ecosystem incorporated into biomass in a particular trophic level. **Biomass** is the total mass, in a unit area at the time of measurement, of living or previously living organisms within a trophic level. Ecosystems have characteristic amounts of biomass at each trophic level. For example, in the English Channel ecosystem the primary producers account for a biomass of 4 g/m<sup>2</sup> (grams per meter squared), while the primary consumers exhibit a biomass of 21 g/m<sup>2</sup>.

The productivity of the primary producers is especially important in any ecosystem because these organisms bring energy to other living organisms by photoautotrophy or chemoautotrophy. The rate at which photosynthetic primary producers incorporate energy from the sun is called **gross primary productivity**. An example of gross primary productivity is shown in the compartment diagram of energy flow within the Silver Springs aquatic ecosystem as shown ([link](#)). In this ecosystem, the total energy accumulated by the primary producers (gross primary productivity) was shown to be 20,810 kcal/m<sup>2</sup>/yr.

Because all organisms need to use some of this energy for their own functions (like respiration and resulting metabolic heat loss) scientists often



refer to the net primary productivity of an ecosystem. **Net primary productivity** is the energy that remains in the primary producers after accounting for the organisms' respiration and heat loss. The net productivity is then available to the primary consumers at the next trophic level. In our Silver Spring example, 13,187 of the 20,810 kcal/m<sup>2</sup>/yr were used for respiration or were lost as heat, leaving 7,632 kcal/m<sup>2</sup>/yr of energy for use by the primary consumers.

## **Ecological Efficiency: The Transfer of Energy between Trophic Levels**

As illustrated in [\[link\]](#), large amounts of energy are lost from the ecosystem from one trophic level to the next level as energy flows from the primary producers through the various trophic levels of consumers and decomposers. The main reason for this loss is the second law of thermodynamics, which states that whenever energy is converted from one form to another, there is a tendency toward disorder (entropy) in the system. In biologic systems, this means a great deal of energy is lost as metabolic heat when the organisms from one trophic level consume the next level. In the Silver Springs ecosystem example ([\[link\]](#)), we see that the primary consumers produced 1103 kcal/m<sup>2</sup>/yr from the 7618 kcal/m<sup>2</sup>/yr of energy available to them from the primary producers. The measurement of energy transfer efficiency between two successive trophic levels is termed the **trophic level transfer efficiency (TLTE)** and is defined by the formula:  
**Equation:**

$$\text{TLTE} = \frac{\text{production at present trophic level}}{\text{production at previous trophic level}} \times 100$$

In Silver Springs, the TLTE between the first two trophic levels was approximately 14.8 percent. The low efficiency of energy transfer between trophic levels is usually the major factor that limits the length of food chains observed in a food web. The fact is, after four to six energy transfers, there is not enough energy left to support another trophic level. In the Lake Ontario example shown in [\[link\]](#), only three energy transfers occurred

between the primary producer, (green algae), and the apex consumer (Chinook salmon).

Ecologists have many different methods of measuring energy transfers within ecosystems. Some transfers are easier or more difficult to measure depending on the complexity of the ecosystem and how much access scientists have to observe the ecosystem. In other words, some ecosystems are more difficult to study than others, and sometimes the quantification of energy transfers has to be estimated.

Another main parameter that is important in characterizing energy flow within an ecosystem is the net production efficiency. **Net production efficiency (NPE)** allows ecologists to quantify how efficiently organisms of a particular trophic level incorporate the energy they receive into biomass; it is calculated using the following formula:

**Equation:**

$$\text{NPE} = \frac{\text{net consumer productivity}}{\text{assimilation}} \times 100$$

**Net consumer productivity** is the energy content available to the organisms of the next trophic level. **Assimilation** is the biomass (energy content generated per unit area) of the present trophic level after accounting for the energy lost due to incomplete ingestion of food, energy used for respiration, and energy lost as waste. Incomplete ingestion refers to the fact that some consumers eat only a part of their food. For example, when a lion kills an antelope, it will eat everything except the hide and bones. The lion is missing the energy-rich bone marrow inside the bone, so the lion does not make use of all the calories its prey could provide.

Thus, NPE measures how efficiently each trophic level uses and incorporates the energy from its food into biomass to fuel the next trophic level. In general, cold-blooded animals (ectotherms), such as invertebrates, fish, amphibians, and reptiles, use less of the energy they obtain for respiration and heat than warm-blooded animals (endotherms), such as birds and mammals. The extra heat generated in endotherms, although an advantage in terms of the activity of these organisms in colder

environments, is a major disadvantage in terms of NPE. Therefore, many endotherms have to eat more often than ectotherms to get the energy they need for survival. In general, NPE for ectotherms is an order of magnitude (10x) higher than for endotherms. For example, the NPE for a caterpillar eating leaves has been measured at 18 percent, whereas the NPE for a squirrel eating acorns may be as low as 1.6 percent.

The inefficiency of energy use by warm-blooded animals has broad implications for the world's food supply. It is widely accepted that the meat industry uses large amounts of crops to feed livestock, and because the NPE is low, much of the energy from animal feed is lost. For example, it costs about 1¢ to produce 1000 dietary calories (kcal) of corn or soybeans, but approximately \$0.19 to produce a similar number of calories growing cattle for beef consumption. The same energy content of milk from cattle is also costly, at approximately \$0.16 per 1000 kcal. Much of this difference is due to the low NPE of cattle. Thus, there has been a growing movement worldwide to promote the consumption of non-meat and non-dairy foods so that less energy is wasted feeding animals for the meat industry.

## **Modeling Ecosystems Energy Flow: Ecological Pyramids**

The structure of ecosystems can be visualized with ecological pyramids, which were first described by the pioneering studies of Charles Elton in the 1920s. **Ecological pyramids** show the relative amounts of various parameters (such as number of organisms, energy, and biomass) across trophic levels.

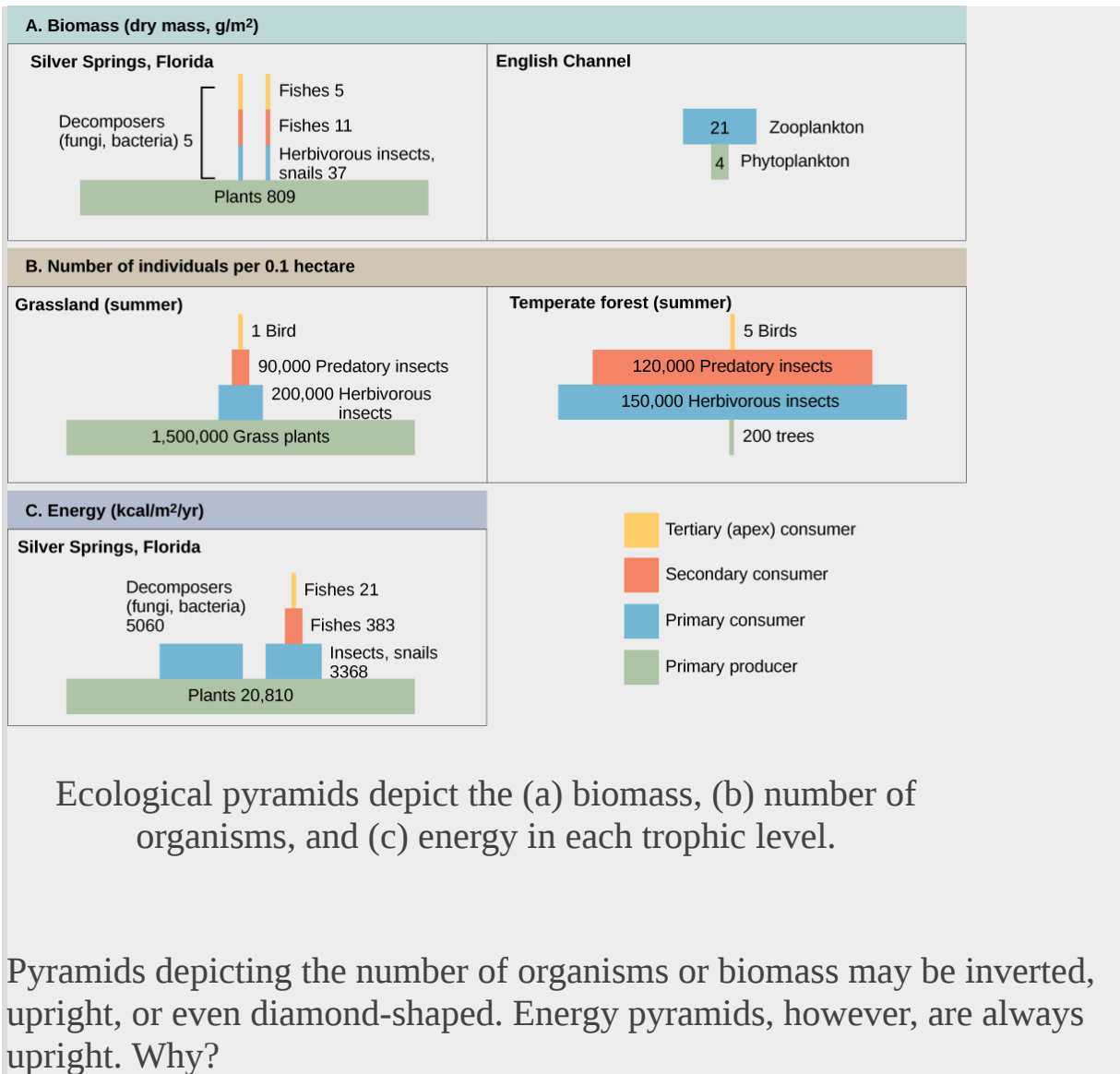
Pyramids of numbers can be either upright or inverted, depending on the ecosystem. As shown in [\[link\]](#), typical grassland during the summer has a base of many plants and the numbers of organisms decrease at each trophic level. However, during the summer in a temperate forest, the base of the pyramid consists of few trees compared with the number of primary consumers, mostly insects. Because trees are large, they have great photosynthetic capability, and dominate other plants in this ecosystem to obtain sunlight. Even in smaller numbers, primary producers in forests are still capable of supporting other trophic levels.

Another way to visualize ecosystem structure is with pyramids of biomass. This pyramid measures the amount of energy converted into living tissue at the different trophic levels. Using the Silver Springs ecosystem example, this data exhibits an upright biomass pyramid ([\[link\]](#)), whereas the pyramid from the English Channel example is inverted. The plants (primary producers) of the Silver Springs ecosystem make up a large percentage of the biomass found there. However, the phytoplankton in the English Channel example make up less biomass than the primary consumers, the zooplankton. As with inverted pyramids of numbers, this inverted pyramid is not due to a lack of productivity from the primary producers, but results from the high turnover rate of the phytoplankton. The phytoplankton are consumed rapidly by the primary consumers, thus, minimizing their biomass at any particular point in time. However, phytoplankton reproduce quickly, thus they are able to support the rest of the ecosystem.

Pyramid ecosystem modeling can also be used to show energy flow through the trophic levels. Notice that these numbers are the same as those used in the energy flow compartment diagram in [\[link\]](#). Pyramids of energy are always upright, and an ecosystem without sufficient primary productivity cannot be supported. All types of ecological pyramids are useful for characterizing ecosystem structure. However, in the study of energy flow through the ecosystem, pyramids of energy are the most consistent and representative models of ecosystem structure ([\[link\]](#)).

**Note:**

Art Connection

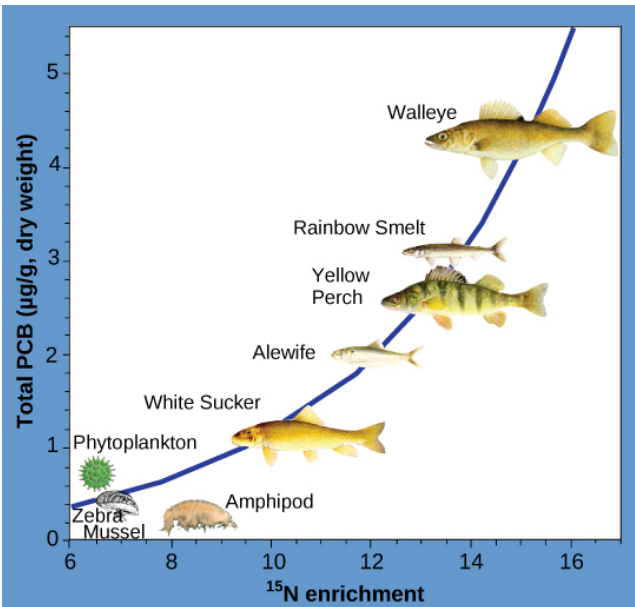


## Consequences of Food Webs: Biological Magnification

One of the most important environmental consequences of ecosystem dynamics is biomagnification. **Biomagnification** is the increasing concentration of persistent, toxic substances in organisms at each trophic level, from the primary producers to the apex consumers. Many substances have been shown to bioaccumulate, including classical studies with the pesticide **dichlorodiphenyltrichloroethane (DDT)**, which was published in the 1960s bestseller, *Silent Spring*, by Rachel Carson. DDT was a

commonly used pesticide before its dangers became known. In some aquatic ecosystems, organisms from each trophic level consumed many organisms of the lower level, which caused DDT to increase in birds (apex consumers) that ate fish. Thus, the birds accumulated sufficient amounts of DDT to cause fragility in their eggshells. This effect increased egg breakage during nesting and was shown to have adverse effects on these bird populations. The use of DDT was banned in the United States in the 1970s.

Other substances that biomagnify are polychlorinated biphenyls (PCBs), which were used in coolant liquids in the United States until their use was banned in 1979, and heavy metals, such as mercury, lead, and cadmium. These substances were best studied in aquatic ecosystems, where fish species at different trophic levels accumulate toxic substances brought through the ecosystem by the primary producers. As illustrated in a study performed by the National Oceanic and Atmospheric Administration (NOAA) in the Saginaw Bay of Lake Huron ([\[link\]](#)), PCB concentrations increased from the ecosystem's primary producers (phytoplankton) through the different trophic levels of fish species. The apex consumer (walleye) has more than four times the amount of PCBs compared to phytoplankton. Also, based on results from other studies, birds that eat these fish may have PCB levels at least one order of magnitude higher than those found in the lake fish.



This chart shows the PCB concentrations found at the various trophic levels in the Saginaw Bay ecosystem of Lake Huron. Numbers on the x-axis reflect enrichment with heavy isotopes of nitrogen ( $^{15}\text{N}$ ), which is a marker for increasing trophic level. Notice that the fish in the higher trophic levels accumulate more PCBs than those in lower trophic levels. (credit: Patricia Van Hoof, NOAA, GLERL)

Other concerns have been raised by the accumulation of heavy metals, such as mercury and cadmium, in certain types of seafood. The United States Environmental Protection Agency (EPA) recommends that pregnant women and young children should not consume any swordfish, shark, king mackerel, or tilefish because of their high mercury content. These individuals are advised to eat fish low in mercury: salmon, tilapia, shrimp,

pollock, and catfish. Biomagnification is a good example of how ecosystem dynamics can affect our everyday lives, even influencing the food we eat.

## Section Summary

Organisms in an ecosystem acquire energy in a variety of ways, which is transferred between trophic levels as the energy flows from the bottom to the top of the food web, with energy being lost at each transfer. The efficiency of these transfers is important for understanding the different behaviors and eating habits of warm-blooded versus cold-blooded animals. Modeling of ecosystem energy is best done with ecological pyramids of energy, although other ecological pyramids provide other vital information about ecosystem structure.

## Art Connections

### Exercise:

#### Problem:

[\[link\]](#) Pyramids depicting the number of organisms or biomass may be inverted, upright, or even diamond-shaped. Energy pyramids, however, are always upright. Why?

---

#### Solution:

[\[link\]](#) Pyramids of organisms may be inverted or diamond-shaped because a large organism, such as a tree, can sustain many smaller organisms. Likewise, a low biomass of organisms can sustain a larger biomass at the next trophic level because the organisms reproduce rapidly and thus supply continuous nourishment. Energy pyramids, however, must always be upright because of the laws of thermodynamics. The first law of thermodynamics states that energy can neither be created nor destroyed; thus, each trophic level must acquire energy from the trophic level below. The second law of thermodynamics states that, during the transfer of energy, some energy



is always lost as heat; thus, less energy is available at each higher trophic level.

## Review Questions

### Exercise:

#### Problem:

The weight of living organisms in an ecosystem at a particular point in time is called:

- a. energy
- b. production
- c. entropy
- d. biomass

---

#### Solution:

D

### Exercise:

#### Problem:

Which term describes the process whereby toxic substances increase along trophic levels of an ecosystem?

- a. biomassification
- b. biomagnification
- c. bioentropy
- d. heterotrophy

---

#### Solution:

B

### Exercise:

**Problem:**

Organisms that can make their own food using inorganic molecules are called:

- a. autotrophs
- b. heterotrophs
- c. photoautotrophs
- d. chemoautotrophs

---

**Solution:**

D

**Exercise:****Problem:**

In the English Channel ecosystem, the number of primary producers is smaller than the number of primary consumers because\_\_\_\_\_.

- a. the apex consumers have a low turnover rate
- b. the primary producers have a low turnover rate
- c. the primary producers have a high turnover rate
- d. the primary consumers have a high turnover rate

---

**Solution:**

C

**Exercise:****Problem:**

What law of chemistry determines how much energy can be transferred when it is converted from one form to another?

- a. the first law of thermodynamics

- b. the second law of thermodynamics
- c. the conservation of matter
- d. the conservation of energy

---

**Solution:**

B

**Free Response**

**Exercise:**

**Problem:**

Compare the three types of ecological pyramids and how well they describe ecosystem structure. Identify which ones can be inverted and give an example of an inverted pyramid for each.

---

**Solution:**

Pyramids of numbers display the number of individual organisms on each trophic level. These pyramids can be either upright or inverted, depending on the number of the organisms. Pyramids of biomass display the weight of organisms at each level. Inverted pyramids of biomass can occur when the primary producer has a high turnover rate. Pyramids of energy are usually upright and are the best representation of energy flow and ecosystem structure.

**Exercise:**

**Problem:**

How does the amount of food a warm blooded-animal (endotherm) eats relate to its net production efficiency (NPE)?

---

**Solution:**

NPE measures the rate at which one trophic level can use and make biomass from what it attained in the previous level, taking into account respiration, defecation, and heat loss. Endotherms have high metabolism and generate a lot of body heat. Although this gives them advantages in their activity level in colder temperatures, these organisms are 10 times less efficient at harnessing the energy from the food they eat compared with cold-blooded animals, and thus have to eat more and more often.

## **Glossary**

### **assimilation**

biomass consumed and assimilated from the previous trophic level after accounting for the energy lost due to incomplete ingestion of food, energy used for respiration, and energy lost as waste

### **biomagnification**

increasing concentrations of persistent, toxic substances in organisms at each trophic level, from the primary producers to the apex consumers

### **biomass**

total weight, at the time of measurement, of living or previously living organisms in a unit area within a trophic level

### **chemoautotroph**

organism capable of synthesizing its own food using energy from inorganic molecules

### **ecological pyramid**

(also, Eltonian pyramid) graphical representation of different trophic levels in an ecosystem based of organism numbers, biomass, or energy content

### **gross primary productivity**

rate at which photosynthetic primary producers incorporate energy from the sun

net consumer productivity

energy content available to the organisms of the next trophic level

net primary productivity

energy that remains in the primary producers after accounting for the organisms' respiration and heat loss

net production efficiency (NPE)

measure of the ability of a trophic level to convert the energy it receives from the previous trophic level into biomass

trophic level transfer efficiency (TLTE)

energy transfer efficiency between two successive trophic levels

## Biogeochemical Cycles

By the end of this section, you will be able to:

- Discuss the biogeochemical cycles of water, carbon, nitrogen, phosphorus, and sulfur
- Explain how human activities have impacted these cycles and the potential consequences for Earth

Energy flows directionally through ecosystems, entering as sunlight (or inorganic molecules for chemoautotrophs) and leaving as heat during the many transfers between trophic levels. However, the matter that makes up living organisms is conserved and recycled. The six most common elements associated with organic molecules—carbon, nitrogen, hydrogen, oxygen, phosphorus, and sulfur—take a variety of chemical forms and may exist for long periods in the atmosphere, on land, in water, or beneath the Earth's surface. Geologic processes, such as weathering, erosion, water drainage, and the subduction of the continental plates, all play a role in this recycling of materials. Because geology and chemistry have major roles in the study of this process, the recycling of inorganic matter between living organisms and their environment is called a **biogeochemical cycle**.

Water contains hydrogen and oxygen, which is essential to all living processes. The **hydrosphere** is the area of the Earth where water movement and storage occurs: as liquid water on the surface and beneath the surface or frozen (rivers, lakes, oceans, groundwater, polar ice caps, and glaciers), and as water vapor in the atmosphere. Carbon is found in all organic macromolecules and is an important constituent of fossil fuels. Nitrogen is a major component of our nucleic acids and proteins and is critical to human agriculture. Phosphorus, a major component of nucleic acid (along with nitrogen), is one of the main ingredients in artificial fertilizers used in agriculture and their associated environmental impacts on our surface water. Sulfur, critical to the 3-D folding of proteins (as in disulfide binding), is released into the atmosphere by the burning of fossil fuels, such as coal.

The cycling of these elements is interconnected. For example, the movement of water is critical for the leaching of nitrogen and phosphate into rivers, lakes, and oceans. Furthermore, the ocean itself is a major reservoir for carbon. Thus, mineral nutrients are cycled, either rapidly or

slowly, through the entire biosphere, from one living organism to another, and between the biotic and abiotic world.

**Note:**

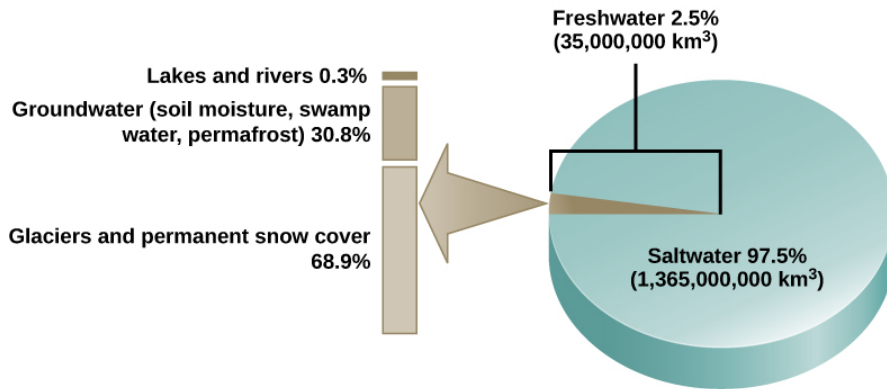
Link to Learning



Head to this [website](#) to learn more about biogeochemical cycles.

## The Water (Hydrologic) Cycle

Water is the basis of all living processes. The human body is more than 1/2 water and human cells are more than 70 percent water. Thus, most land animals need a supply of fresh water to survive. However, when examining the stores of water on Earth, 97.5 percent of it is non-potable salt water ([link](#)). Of the remaining water, 99 percent is locked underground as water or as ice. Thus, less than 1 percent of fresh water is easily accessible from lakes and rivers. Many living things, such as plants, animals, and fungi, are dependent on the small amount of fresh surface water supply, a lack of which can have massive effects on ecosystem dynamics. Humans, of course, have developed technologies to increase water availability, such as digging wells to harvest groundwater, storing rainwater, and using desalination to obtain drinkable water from the ocean. Although this pursuit of drinkable water has been ongoing throughout human history, the supply of fresh water is still a major issue in modern times.



Only 2.5 percent of water on Earth is fresh water, and less than 1 percent of fresh water is easily accessible to living things.

Water cycling is extremely important to ecosystem dynamics. Water has a major influence on climate and, thus, on the environments of ecosystems, some located on distant parts of the Earth. Most of the water on Earth is stored for long periods in the oceans, underground, and as ice. [\[link\]](#) illustrates the average time that an individual water molecule may spend in the Earth's major water reservoirs. **Residence time** is a measure of the average time an individual water molecule stays in a particular reservoir. A large amount of the Earth's water is locked in place in these reservoirs as ice, beneath the ground, and in the ocean, and, thus, is unavailable for short-term cycling (only surface water can evaporate).



Average Residence Time for Water Molecules	
Biospheric (in living organisms)	1 week
Atmospheric	1.5 weeks
Rivers	2 weeks
Soil moisture	2 weeks–1 year
Swamps	1–10 years
Lakes & reservoirs	10 years
Oceans & seas	4,000 years
Groundwater	2 weeks to 10,000 years
Glaciers and permafrost	1,000–10,000 years

This graph shows the average residence time for water molecules in the Earth's water reservoirs.

There are various processes that occur during the cycling of water, shown in [\[link\]](#). These processes include the following:

- evaporation/sublimation
- condensation/precipitation
- subsurface water flow
- surface runoff/snowmelt
- streamflow

The water cycle is driven by the sun's energy as it warms the oceans and other surface waters. This leads to the evaporation (water to water vapor) of liquid surface water and the sublimation (ice to water vapor) of frozen water, which deposits large amounts of water vapor into the atmosphere. Over time, this water vapor condenses into clouds as liquid or frozen droplets and is eventually followed by precipitation (rain or snow), which returns water to the Earth's surface. Rain eventually permeates into the ground, where it may evaporate again if it is near the surface, flow beneath

the surface, or be stored for long periods. More easily observed is surface runoff: the flow of fresh water either from rain or melting ice. Runoff can then make its way through streams and lakes to the oceans or flow directly to the oceans themselves.

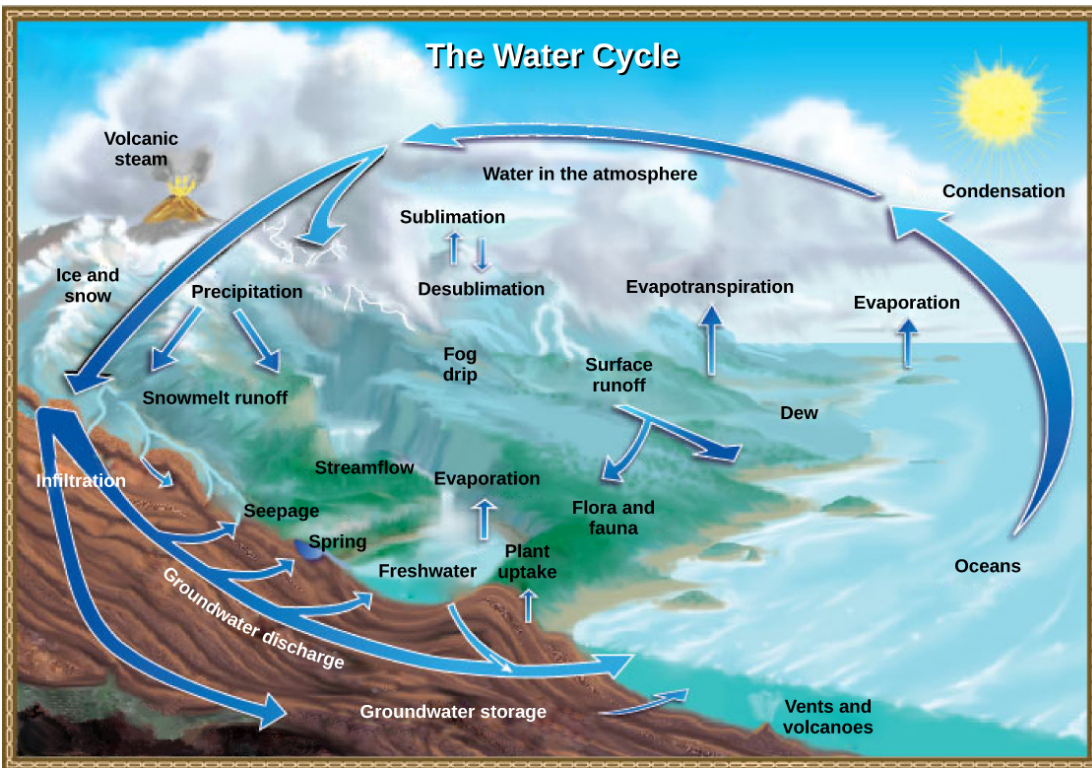
**Note:**

Link to Learning



Head to this [website](#) to learn more about the world's fresh water supply.

Rain and surface runoff are major ways in which minerals, including carbon, nitrogen, phosphorus, and sulfur, are cycled from land to water. The environmental effects of runoff will be discussed later as these cycles are described.



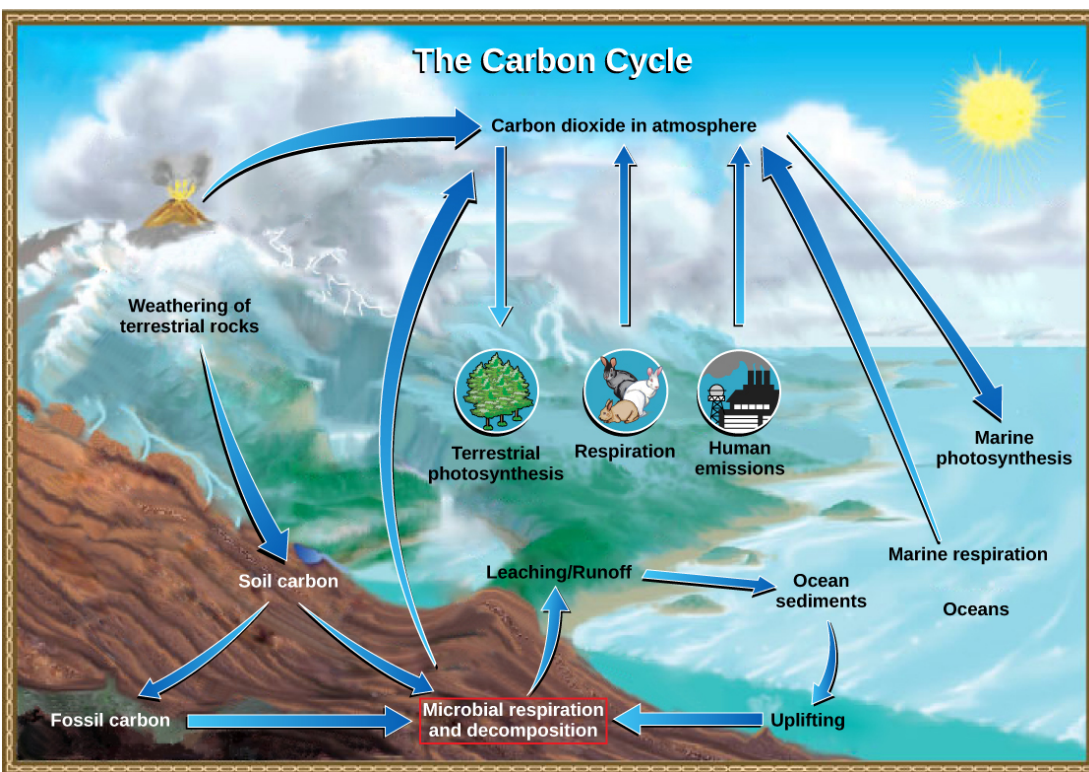
Water from the land and oceans enters the atmosphere by evaporation or sublimation, where it condenses into clouds and falls as rain or snow. Precipitated water may enter freshwater bodies or infiltrate the soil. The cycle is complete when surface or groundwater reenters the ocean. (credit: modification of work by John M. Evans and Howard Perlman, USGS)

## The Carbon Cycle

Carbon is the second most abundant element in living organisms. Carbon is present in all organic molecules, and its role in the structure of macromolecules is of primary importance to living organisms. Carbon compounds contain especially high energy, particularly those derived from fossilized organisms, mainly plants, which humans use as fuel. Since the 1800s, the number of countries using massive amounts of fossil fuels has increased. Since the beginning of the Industrial Revolution, global demand for the Earth's limited fossil fuel supplies has risen; therefore, the amount of

carbon dioxide in our atmosphere has increased. This increase in carbon dioxide has been associated with climate change and other disturbances of the Earth's ecosystems and is a major environmental concern worldwide. Thus, the “carbon footprint” is based on how much carbon dioxide is produced and how much fossil fuel countries consume.

The carbon cycle is most easily studied as two interconnected sub-cycles: one dealing with rapid carbon exchange among living organisms and the other dealing with the long-term cycling of carbon through geologic processes. The entire carbon cycle is shown in [\[link\]](#).



Carbon dioxide gas exists in the atmosphere and is dissolved in water. Photosynthesis converts carbon dioxide gas to organic carbon, and respiration cycles the organic carbon back into carbon dioxide gas. Long-term storage of organic carbon occurs when matter from living organisms is buried deep underground and becomes fossilized. Volcanic activity and, more recently, human emissions, bring this stored carbon back

into the carbon cycle. (credit: modification of work by John M. Evans and Howard Perlman, USGS)

**Note:**

Link to Learning



Click this [link](#) to read information about the United States Carbon Cycle Science Program.

## The Biological Carbon Cycle

Living organisms are connected in many ways, even between ecosystems. A good example of this connection is the exchange of carbon between autotrophs and heterotrophs within and between ecosystems by way of atmospheric carbon dioxide. Carbon dioxide is the basic building block that most autotrophs use to build multi-carbon, high energy compounds, such as glucose. The energy harnessed from the sun is used by these organisms to form the covalent bonds that link carbon atoms together. These chemical bonds thereby store this energy for later use in the process of respiration. Most terrestrial autotrophs obtain their carbon dioxide directly from the atmosphere, while marine autotrophs acquire it in the dissolved form (carbonic acid,  $\text{H}_2\text{CO}_3^-$ ). However carbon dioxide is acquired, a by-product of the process is oxygen. The photosynthetic organisms are responsible for depositing approximately 21 percent oxygen content of the atmosphere that we observe today.

Heterotrophs and autotrophs are partners in biological carbon exchange (especially the primary consumers, largely herbivores). Heterotrophs acquire the high-energy carbon compounds from the autotrophs by consuming them, and breaking them down by respiration to obtain cellular energy, such as ATP. The most efficient type of respiration, aerobic respiration, requires oxygen obtained from the atmosphere or dissolved in water. Thus, there is a constant exchange of oxygen and carbon dioxide between the autotrophs (which need the carbon) and the heterotrophs (which need the oxygen). Gas exchange through the atmosphere and water is one way that the carbon cycle connects all living organisms on Earth.

## The Biogeochemical Carbon Cycle

The movement of carbon through the land, water, and air is complex, and in many cases, it occurs much more slowly geologically than as seen between living organisms. Carbon is stored for long periods in what are known as carbon reservoirs, which include the atmosphere, bodies of liquid water (mostly oceans), ocean sediment, soil, land sediments (including fossil fuels), and the Earth's interior.

As stated, the atmosphere is a major reservoir of carbon in the form of carbon dioxide and is essential to the process of photosynthesis. The level of carbon dioxide in the atmosphere is greatly influenced by the reservoir of carbon in the oceans. The exchange of carbon between the atmosphere and water reservoirs influences how much carbon is found in each location, and each one affects the other reciprocally. Carbon dioxide ( $\text{CO}_2$ ) from the atmosphere dissolves in water and combines with water molecules to form carbonic acid, and then it ionizes to carbonate and bicarbonate ions ([\[link\]](#))

Step 1:  $\text{CO}_2$  (atmospheric)  $\rightleftharpoons$   $\text{CO}_2$  (dissolved)

Step 2:  $\text{CO}_2$  (dissolved) +  $\text{H}_2\text{O}$   $\rightleftharpoons$   $\text{H}_2\text{CO}_3$  (carbonic acid)

Step 3:  $\text{H}_2\text{CO}_3$   $\rightleftharpoons$   $\text{H}^+$  +  $\text{HCO}_3^-$  (bicarbonate ion)

Step 4:  $\text{HCO}_3^-$   $\rightleftharpoons$   $\text{H}^+$  +  $\text{CO}_3^{2-}$  (carbonate ion)



Carbon dioxide reacts with water to form bicarbonate and carbonate ions.

The equilibrium coefficients are such that more than 90 percent of the carbon in the ocean is found as bicarbonate ions. Some of these ions combine with seawater calcium to form calcium carbonate ( $\text{CaCO}_3$ ), a major component of marine organism shells. These organisms eventually form sediments on the ocean floor. Over geologic time, the calcium carbonate forms limestone, which comprises the largest carbon reservoir on Earth.

On land, carbon is stored in soil as a result of the decomposition of living organisms (by decomposers) or from weathering of terrestrial rock and minerals. This carbon can be leached into the water reservoirs by surface runoff. Deeper underground, on land and at sea, are fossil fuels: the anaerobically decomposed remains of plants that take millions of years to form. Fossil fuels are considered a non-renewable resource because their use far exceeds their rate of formation. A **non-renewable resource**, such as fossil fuel, is either regenerated very slowly or not at all. Another way for carbon to enter the atmosphere is from land (including land beneath the surface of the ocean) by the eruption of volcanoes and other geothermal systems. Carbon sediments from the ocean floor are taken deep within the Earth by the process of **subduction**: the movement of one tectonic plate beneath another. Carbon is released as carbon dioxide when a volcano erupts or from volcanic hydrothermal vents.

Carbon dioxide is also added to the atmosphere by the animal husbandry practices of humans. The large numbers of land animals raised to feed the Earth's growing population results in increased carbon dioxide levels in the atmosphere due to farming practices and the respiration and methane production. This is another example of how human activity indirectly affects biogeochemical cycles in a significant way. Although much of the debate about the future effects of increasing atmospheric carbon on climate change focuses on fossil fuels, scientists take natural processes, such as

volcanoes and respiration, into account as they model and predict the future impact of this increase.

## The Nitrogen Cycle

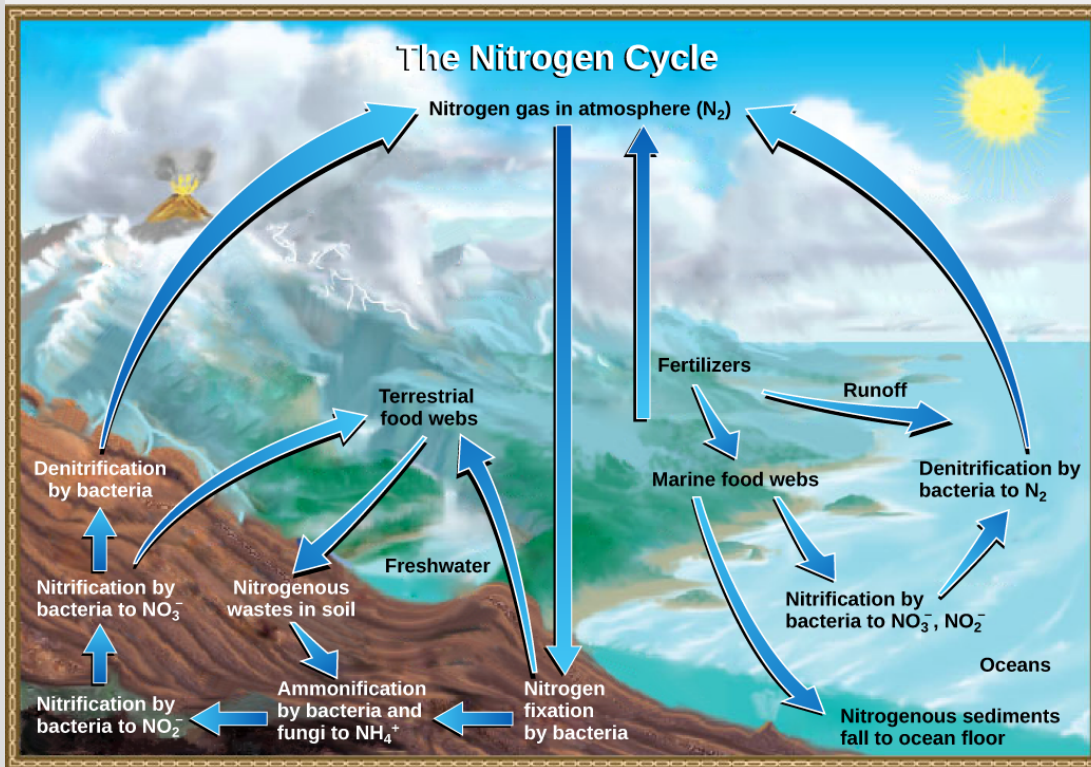
Getting nitrogen into the living world is difficult. Plants and phytoplankton are not equipped to incorporate nitrogen from the atmosphere (which exists as tightly bonded, triple covalent  $N_2$ ) even though this molecule comprises approximately 78 percent of the atmosphere. Nitrogen enters the living world via free-living and symbiotic bacteria, which incorporate nitrogen into their macromolecules through nitrogen fixation (conversion of  $N_2$ ). Cyanobacteria live in most aquatic ecosystems where sunlight is present; they play a key role in nitrogen fixation. Cyanobacteria are able to use inorganic sources of nitrogen to “fix” nitrogen. *Rhizobium* bacteria live symbiotically in the root nodules of legumes (such as peas, beans, and peanuts) and provide them with the organic nitrogen they need. Free-living bacteria, such as *Azotobacter*, are also important nitrogen fixers.

Organic nitrogen is especially important to the study of ecosystem dynamics since many ecosystem processes, such as primary production and decomposition, are limited by the available supply of nitrogen. As shown in [\[link\]](#), the nitrogen that enters living systems by nitrogen fixation is successively converted from organic nitrogen back into nitrogen gas by bacteria. This process occurs in three steps in terrestrial systems: ammonification, nitrification, and denitrification. First, the ammonification process converts nitrogenous waste from living animals or from the remains of dead animals into ammonium ( $NH_4^+$ ) by certain bacteria and fungi. Second, the ammonium is converted to nitrites ( $NO_2^-$ ) by nitrifying bacteria, such as *Nitrosomonas*, through nitrification. Subsequently, nitrites are converted to nitrates ( $NO_3^-$ ) by similar organisms. Third, the process of denitrification occurs, whereby bacteria, such as *Pseudomonas* and *Clostridium*, convert the nitrates into nitrogen gas, allowing it to re-enter the atmosphere.

**Note:**



## Art Connection



Nitrogen enters the living world from the atmosphere via nitrogen-fixing bacteria. This nitrogen and nitrogenous waste from animals is then processed back into gaseous nitrogen by soil bacteria, which also supply terrestrial food webs with the organic nitrogen they need. (credit: modification of work by John M. Evans and Howard Perlman, USGS)

Which of the following statements about the nitrogen cycle is false?

- a. Ammonification converts organic nitrogenous matter from living organisms into ammonium ( $NH_4^+$ ).
- b. Denitrification by bacteria converts nitrates ( $NO_3^-$ ) to nitrogen gas ( $N_2$ ).
- c. Nitrification by bacteria converts nitrates ( $NO_3^-$ ) to nitrites ( $NO_2^-$ ).
- d. Nitrogen fixing bacteria convert nitrogen gas ( $N_2$ ) into organic compounds.

Human activity can release nitrogen into the environment by two primary means: the combustion of fossil fuels, which releases different nitrogen oxides, and by the use of artificial fertilizers in agriculture, which are then washed into lakes, streams, and rivers by surface runoff. Atmospheric nitrogen is associated with several effects on Earth's ecosystems including the production of acid rain (as nitric acid,  $\text{HNO}_3$ ) and greenhouse gas (as nitrous oxide,  $\text{N}_2\text{O}$ ) potentially causing climate change. A major effect from fertilizer runoff is saltwater and freshwater **eutrophication**, a process whereby nutrient runoff causes the excess growth of microorganisms, depleting dissolved oxygen levels and killing ecosystem fauna.

A similar process occurs in the marine nitrogen cycle, where the ammonification, nitrification, and denitrification processes are performed by marine bacteria. Some of this nitrogen falls to the ocean floor as sediment, which can then be moved to land in geologic time by uplift of the Earth's surface and thereby incorporated into terrestrial rock. Although the movement of nitrogen from rock directly into living systems has been traditionally seen as insignificant compared with nitrogen fixed from the atmosphere, a recent study showed that this process may indeed be significant and should be included in any study of the global nitrogen cycle. [\[footnote\]](#)

Scott L. Morford, Benjamin Z. Houlton, and Randy A. Dahlgren, "Increased Forest Ecosystem Carbon and Nitrogen Storage from Nitrogen Rich Bedrock," *Nature* 477, no. 7362 (2011): 78–81.

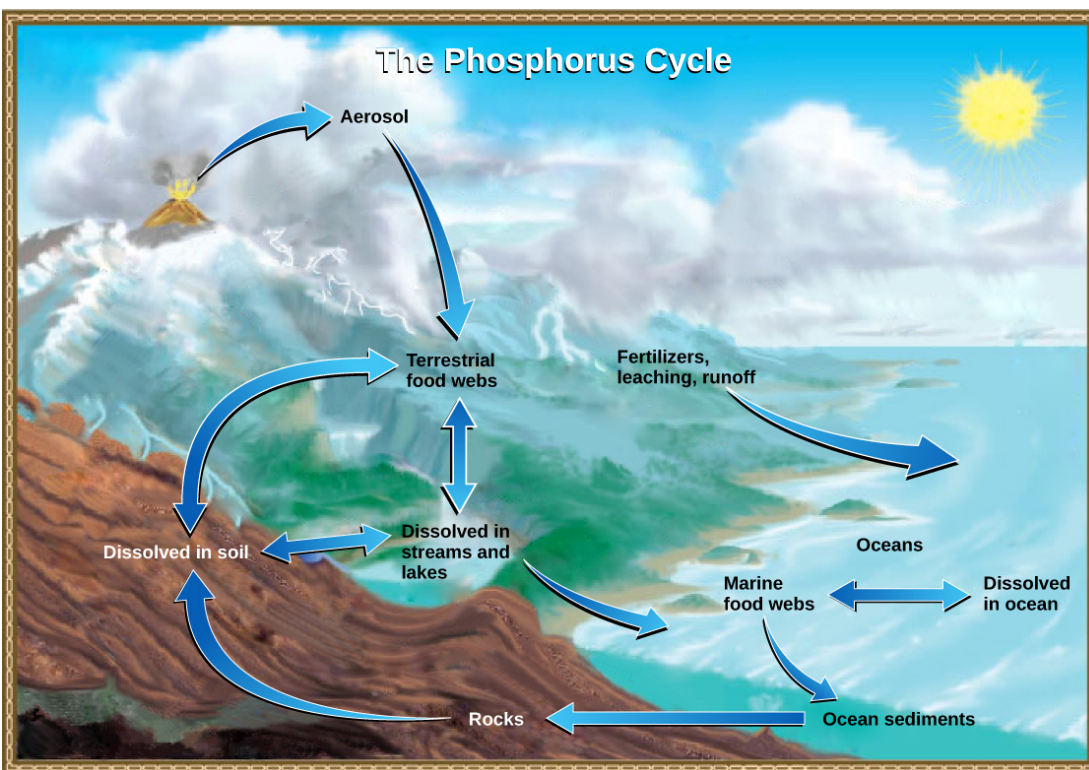
## The Phosphorus Cycle

Phosphorus is an essential nutrient for living processes; it is a major component of nucleic acid and phospholipids, and, as calcium phosphate, makes up the supportive components of our bones. Phosphorus is often the limiting nutrient (necessary for growth) in aquatic ecosystems ([\[link\]](#)).

Phosphorus occurs in nature as the phosphate ion ( $\text{PO}_4^{3-}$ ). In addition to phosphate runoff as a result of human activity, natural surface runoff occurs when it is leached from phosphate-containing rock by weathering, thus sending phosphates into rivers, lakes, and the ocean. This rock has its origins in the ocean. Phosphate-containing ocean sediments form primarily

from the bodies of ocean organisms and from their excretions. However, in remote regions, volcanic ash, aerosols, and mineral dust may also be significant phosphate sources. This sediment then is moved to land over geologic time by the uplifting of areas of the Earth's surface.

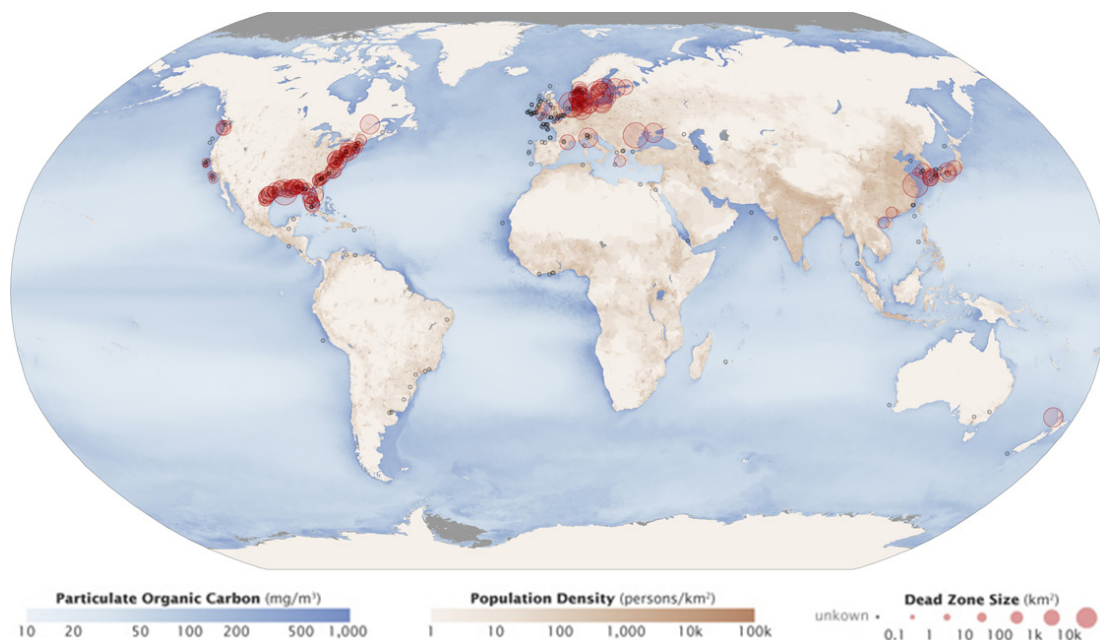
Phosphorus is also reciprocally exchanged between phosphate dissolved in the ocean and marine ecosystems. The movement of phosphate from the ocean to the land and through the soil is extremely slow, with the average phosphate ion having an oceanic residence time between 20,000 and 100,000 years.



In nature, phosphorus exists as the phosphate ion ( $\text{PO}_4^{3-}$ ). Weathering of rocks and volcanic activity releases phosphate into the soil, water, and air, where it becomes available to terrestrial food webs. Phosphate enters the oceans via surface runoff, groundwater flow, and river flow. Phosphate dissolved in ocean water cycles into marine food webs. Some phosphate from the marine food webs falls to the ocean floor, where it

forms sediment. (credit: modification of work by John M. Evans and Howard Perlman, USGS)

Excess phosphorus and nitrogen that enters these ecosystems from fertilizer runoff and from sewage causes excessive growth of microorganisms and depletes the dissolved oxygen, which leads to the death of many ecosystem fauna, such as shellfish and finfish. This process is responsible for dead zones in lakes and at the mouths of many major rivers ([link](#)).



Dead zones occur when phosphorus and nitrogen from fertilizers cause excessive growth of microorganisms, which depletes oxygen and kills fauna. Worldwide, large dead zones are found in coastal areas of high population density. (credit: NASA Earth Observatory)



A **dead zone** is an area within a freshwater or marine ecosystem where large areas are depleted of their normal flora and fauna; these zones can be caused by eutrophication, oil spills, dumping of toxic chemicals, and other human activities. The number of dead zones has been increasing for several years, and more than 400 of these zones were present as of 2008. One of the worst dead zones is off the coast of the United States in the Gulf of Mexico, where fertilizer runoff from the Mississippi River basin has created a dead zone of over 8463 square miles. Phosphate and nitrate runoff from fertilizers also negatively affect several lake and bay ecosystems including the Chesapeake Bay in the eastern United States.

**Note:**

Everyday Connection  
Chesapeake Bay



(a)



(b)

This (a) satellite image shows the Chesapeake Bay, an ecosystem affected by phosphate and nitrate runoff. A (b) member of the Army Corps of Engineers holds a clump of oysters being used as a part of the oyster restoration effort in

the bay. (credit a: modification  
of work by NASA/MODIS;  
credit b: modification of work  
by U.S. Army)

The Chesapeake Bay has long been valued as one of the most scenic areas on Earth; it is now in distress and is recognized as a declining ecosystem. In the 1970s, the Chesapeake Bay was one of the first ecosystems to have identified dead zones, which continue to kill many fish and bottom-dwelling species, such as clams, oysters, and worms. Several species have declined in the Chesapeake Bay due to surface water runoff containing excess nutrients from artificial fertilizer used on land. The source of the fertilizers (with high nitrogen and phosphate content) is not limited to agricultural practices. There are many nearby urban areas and more than 150 rivers and streams empty into the bay that are carrying fertilizer runoff from lawns and gardens. Thus, the decline of the Chesapeake Bay is a complex issue and requires the cooperation of industry, agriculture, and everyday homeowners.

Of particular interest to conservationists is the oyster population; it is estimated that more than 200,000 acres of oyster reefs existed in the bay in the 1700s, but that number has now declined to only 36,000 acres. Oyster harvesting was once a major industry for Chesapeake Bay, but it declined 88 percent between 1982 and 2007. This decline was due not only to fertilizer runoff and dead zones but also to overharvesting. Oysters require a certain minimum population density because they must be in close proximity to reproduce. Human activity has altered the oyster population and locations, greatly disrupting the ecosystem.

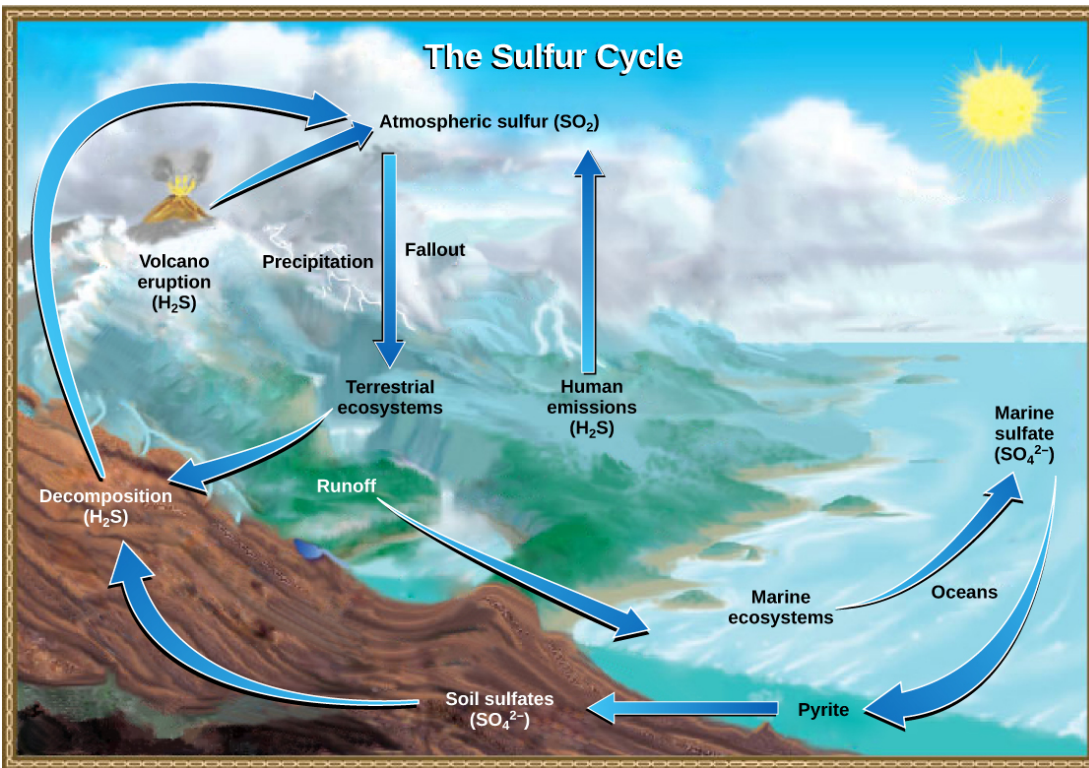
The restoration of the oyster population in the Chesapeake Bay has been ongoing for several years with mixed success. Not only do many people find oysters good to eat, but they also clean up the bay. Oysters are filter feeders, and as they eat, they clean the water around them. In the 1700s, it was estimated that it took only a few days for the oyster population to filter the entire volume of the bay. Today, with changed water conditions, it is estimated that the present population would take nearly a year to do the same job.

Restoration efforts have been ongoing for several years by non-profit organizations, such as the Chesapeake Bay Foundation. The restoration goal is to find a way to increase population density so the oysters can reproduce more efficiently. Many disease-resistant varieties (developed at the Virginia Institute of Marine Science for the College of William and Mary) are now available and have been used in the construction of experimental oyster reefs. Efforts to clean and restore the bay by Virginia and Delaware have been hampered because much of the pollution entering the bay comes from other states, which stresses the need for inter-state cooperation to gain successful restoration.

The new, hearty oyster strains have also spawned a new and economically viable industry—oyster aquaculture—which not only supplies oysters for food and profit, but also has the added benefit of cleaning the bay.

## The Sulfur Cycle

Sulfur is an essential element for the macromolecules of living things. As a part of the amino acid cysteine, it is involved in the formation of disulfide bonds within proteins, which help to determine their 3-D folding patterns, and hence their functions. As shown in [\[link\]](#), sulfur cycles between the oceans, land, and atmosphere. Atmospheric sulfur is found in the form of sulfur dioxide ( $\text{SO}_2$ ) and enters the atmosphere in three ways: from the decomposition of organic molecules, from volcanic activity and geothermal vents, and from the burning of fossil fuels by humans.



Sulfur dioxide from the atmosphere becomes available to terrestrial and marine ecosystems when it is dissolved in precipitation as weak sulfuric acid or when it falls directly to the Earth as fallout. Weathering of rocks also makes sulfates available to terrestrial ecosystems. Decomposition of living organisms returns sulfates to the ocean, soil and atmosphere. (credit: modification of work by John M. Evans and Howard Perlman, USGS)

On land, sulfur is deposited in four major ways: precipitation, direct fallout from the atmosphere, rock weathering, and geothermal vents ([\[link\]](#)). Atmospheric sulfur is found in the form of sulfur dioxide ( $\text{SO}_2$ ), and as rain falls through the atmosphere, sulfur is dissolved in the form of weak sulfuric acid ( $\text{H}_2\text{SO}_4$ ). Sulfur can also fall directly from the atmosphere in a process called **fallout**. Also, the weathering of sulfur-containing rocks releases sulfur into the soil. These rocks originate from ocean sediments that are moved to land by the geologic uplifting of ocean sediments.



Terrestrial ecosystems can then make use of these soil sulfates ( $\text{SO}_4^-$ ), and upon the death and decomposition of these organisms, release the sulfur back into the atmosphere as hydrogen sulfide ( $\text{H}_2\text{S}$ ) gas.



At this sulfur vent in Lassen Volcanic National Park in northeastern California, the yellowish sulfur deposits are visible near the mouth of the vent.

Sulfur enters the ocean via runoff from land, from atmospheric fallout, and from underwater geothermal vents. Some ecosystems ([link](#)) rely on chemoautotrophs using sulfur as a biological energy source. This sulfur then supports marine ecosystems in the form of sulfates.

Human activities have played a major role in altering the balance of the global sulfur cycle. The burning of large quantities of fossil fuels, especially from coal, releases larger amounts of hydrogen sulfide gas into the atmosphere. As rain falls through this gas, it creates the phenomenon known as acid rain. **Acid rain** is corrosive rain caused by rainwater falling to the ground through sulfur dioxide gas, turning it into weak sulfuric acid,

which causes damage to aquatic ecosystems. Acid rain damages the natural environment by lowering the pH of lakes, which kills many of the resident fauna; it also affects the man-made environment through the chemical degradation of buildings. For example, many marble monuments, such as the Lincoln Memorial in Washington, DC, have suffered significant damage from acid rain over the years. These examples show the wide-ranging effects of human activities on our environment and the challenges that remain for our future.

**Note:**

Link to Learning



Click this [link](#) to learn more about global climate change.

## Section Summary

Mineral nutrients are cycled through ecosystems and their environment. Of particular importance are water, carbon, nitrogen, phosphorus, and sulfur. All of these cycles have major impacts on ecosystem structure and function. As human activities have caused major disturbances to these cycles, their study and modeling is especially important. A variety of human activities, such as pollution, oil spills, and events) have damaged ecosystems, potentially causing global climate change. The health of Earth depends on understanding these cycles and how to protect the environment from irreversible damage.

## Art Connections

**Exercise:****Problem:**

[\[link\]](#) Which of the following statements about the nitrogen cycle is false?

- a. Ammonification converts organic nitrogenous matter from living organisms into ammonium ( $\text{NH}_4^+$ ).
- b. Denitrification by bacteria converts nitrates ( $\text{NO}_3^-$ ) to nitrogen gas ( $\text{N}_2$ ).
- c. Nitrification by bacteria converts nitrates ( $\text{NO}_3^-$ ) to nitrites ( $\text{NO}_2^-$ ).
- d. Nitrogen fixing bacteria convert nitrogen gas ( $\text{N}_2$ ) into organic compounds.

---

**Solution:**

[\[link\]](#) C: Nitrification by bacteria converts nitrates ( $\text{NO}_3^-$ ) to nitrites ( $\text{NO}_2^-$ ).

**Review Questions****Exercise:****Problem:**

The movement of mineral nutrients through organisms and their environment is called a \_\_\_\_\_ cycle.

- a. biological
- b. bioaccumulation
- c. biogeochemical
- d. biochemical

---

**Solution:**

C

**Exercise:**

**Problem:** Carbon is present in the atmosphere as \_\_\_\_\_.

- a. carbon dioxide
  - b. carbonate ion
  - c. carbon dust
  - d. carbon monoxide
- 

**Solution:**

A

**Exercise:**

**Problem:** The majority of water found on Earth is:

- a. ice
  - b. water vapor
  - c. fresh water
  - d. salt water
- 

**Solution:**

D

**Exercise:**

**Problem:**

The average time a molecule spends in its reservoir is known as \_\_\_\_\_.

- a. residence time
- b. restriction time
- c. resilience time

d. storage time

---

**Solution:**

A

**Exercise:**

**Problem:**

The process whereby oxygen is depleted by the growth of microorganisms due to excess nutrients in aquatic systems is called \_\_\_\_\_.

- a. dead zoning
- b. eutrophication
- c. retrofication
- d. depletion

---

**Solution:**

B

**Exercise:**

**Problem:**

The process whereby nitrogen is brought into organic molecules is called \_\_\_\_\_.

- a. nitrification
- b. denitrification
- c. nitrogen fixation
- d. nitrogen cycling

---

**Solution:**

C

## Free Response

### Exercise:

#### Problem:

Describe nitrogen fixation and why it is important to agriculture.

---

#### Solution:

Nitrogen fixation is the process of bringing nitrogen gas from the atmosphere and incorporating it into organic molecules. Most plants do not have this capability and must rely on free-living or symbiotic bacteria to do this. As nitrogen is often the limiting nutrient in the growth of crops, farmers make use of artificial fertilizers to provide a nitrogen source to the plants as they grow.

### Exercise:

#### Problem:

What are the factors that cause dead zones? Describe eutrophication, in particular, as a cause.

---

#### Solution:

Many factors can kill life in a lake or ocean, such as eutrophication by nutrient-rich surface runoff, oil spills, toxic waste spills, changes in climate, and the dumping of garbage into the ocean. Eutrophication is a result of nutrient-rich runoff from land using artificial fertilizers high in nitrogen and phosphorus. These nutrients cause the rapid and excessive growth of microorganisms, which deplete local dissolved oxygen and kill many fish and other aquatic organisms.

### Exercise:

**Problem:**

Why are drinking water supplies still a major concern for many countries?

---

**Solution:**

Most of the water on Earth is salt water, which humans cannot drink unless the salt is removed. Some fresh water is locked in glaciers and polar ice caps, or is present in the atmosphere. The Earth's water supplies are threatened by pollution and exhaustion. The effort to supply fresh drinking water to the planet's ever-expanding human population is seen as a major challenge in this century.

**Glossary****acid rain**

corrosive rain caused by rainwater falling to the ground through sulfur dioxide gas, turning it into weak sulfuric acid; can damage structures and ecosystems

**biogeochemical cycle**

cycling of mineral nutrients through ecosystems and through the non-living world

**dead zone**

area within an ecosystem in lakes and near the mouths of rivers where large areas of ecosystems are depleted of their normal flora and fauna; these zones can be caused by eutrophication, oil spills, dumping of toxic chemicals, and other human activities

**eutrophication**

process whereby nutrient runoff causes the excess growth of microorganisms, depleting dissolved oxygen levels and killing ecosystem fauna

**fallout**

direct deposit of solid minerals on land or in the ocean from the atmosphere

hydrosphere

area of the Earth where water movement and storage occurs

non-renewable resource

resource, such as fossil fuel, that is either regenerated very slowly or not at all

residence time

measure of the average time an individual water molecule stays in a particular reservoir

subduction

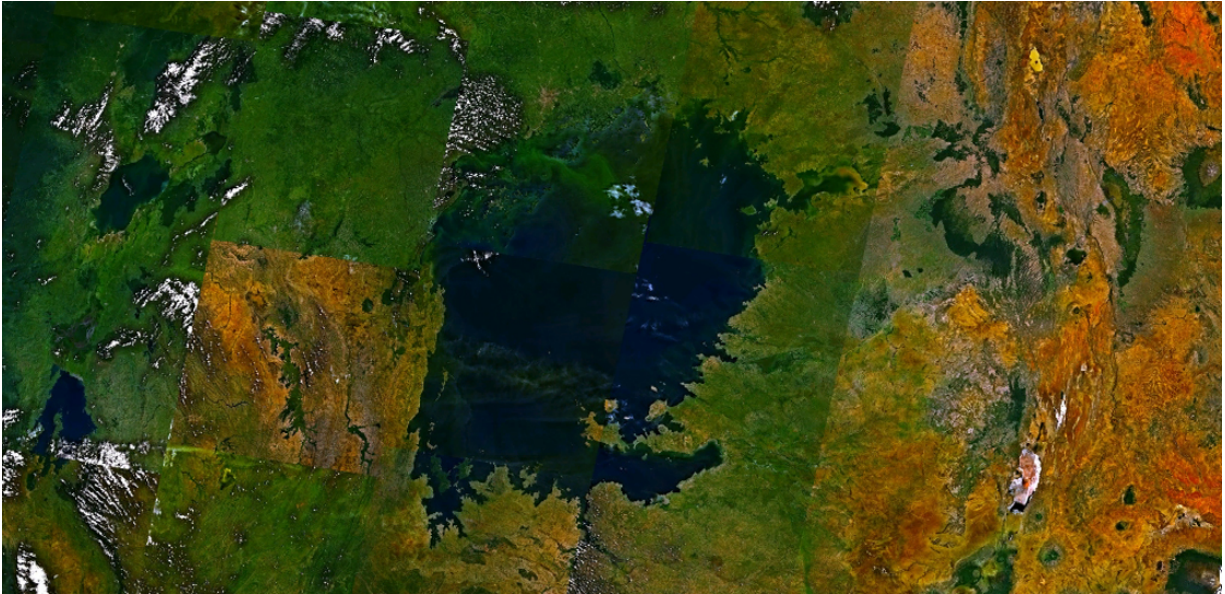
movement of one tectonic plate beneath another



## Introduction

class="introduction"

Lake  
Victoria in  
Africa,  
shown in this  
satellite  
image, was  
the site of  
one of the  
most  
extraordinar  
y  
evolutionary  
findings on  
the planet, as  
well as a  
casualty of  
devastating  
biodiversity  
loss. (credit:  
modification  
of work by  
Rishabh  
Tatiraju,  
using NASA  
World Wind  
software)



In the 1980s, biologists working in Lake Victoria in Africa discovered one of the most extraordinary products of evolution on the planet. Located in the Great Rift Valley, Lake Victoria is a large lake about 68,900 km<sup>2</sup> in area (larger than Lake Huron, the second largest of North America's Great Lakes). Biologists were studying species of a family of fish called cichlids. They found that as they sampled for fish in different locations of the lake, they never stopped finding new species, and they identified nearly 500 evolved types of cichlids. But while studying these variations, they quickly discovered that the invasive Nile Perch was destroying the lake's cichlid population, bringing hundreds of cichlid species to extinction with devastating rapidity.

## The Biodiversity Crisis

By the end of this section, you will be able to:

- Define biodiversity
- Describe biodiversity as the equilibrium of naturally fluctuating rates of extinction and speciation
- Identify historical causes of high extinction rates in Earth's history

Traditionally, ecologists have measured **biodiversity**, a general term for the variety present in the biosphere, by taking into account both the number of species and their commonness. Biodiversity can be estimated at a number of levels of organization of living things. These estimation indexes, which came from information theory, are most useful as a first step in quantifying biodiversity between and within ecosystems; they are less useful when the main concern among conservation biologists is simply the loss of biodiversity. However, biologists recognize that measures of biodiversity, in terms of species diversity, may help focus efforts to preserve the biologically or technologically important elements of biodiversity.

The Lake Victoria cichlids provide an example through which we can begin to understand biodiversity. The biologists studying cichlids in the 1980s discovered hundreds of cichlid species representing a variety of specializations to particular habitat types and specific feeding strategies: eating plankton floating in the water, scraping and then eating algae from rocks, eating insect larvae from the bottom, and eating the eggs of other species of cichlid. The cichlids of Lake Victoria are the product of an **adaptive radiation**. An adaptive radiation is a rapid (less than three million years in the case of the Lake Victoria cichlids) branching through speciation of a phylogenetic tree into many closely related species; typically, the species “radiate” into different habitats and niches. The Galápagos finches are an example of a modest adaptive radiation with 15 species. The cichlids of Lake Victoria are an example of a spectacular adaptive radiation that includes about 500 species.

At the time biologists were making this discovery, some species began to quickly disappear. A culprit in these declines was a species of large fish that was introduced to Lake Victoria by fisheries to feed the people living around the lake. The Nile perch was introduced in 1963, but lay low until the 1980s when its populations began to surge. The Nile perch population grew by consuming cichlids, driving species after species to the point of **extinction** (the disappearance of a species). In fact, there were several factors that played a role in the extinction of perhaps 200 cichlid species in Lake Victoria: the Nile perch, declining lake water quality due to agriculture and land clearing on the shores of Lake Victoria, and increased fishing pressure. Scientists had not even catalogued all of the species present—so many were lost that were never named. The diversity is now a shadow of what it once was.

The cichlids of Lake Victoria are a thumbnail sketch of contemporary rapid species loss that occurs all over Earth and is caused by human activity. Extinction is a natural process of macroevolution that occurs at the rate of about one out of 1 million species becoming extinct per year. The fossil record reveals that there have been five periods of mass extinction in history with much higher rates of species loss, and the rate of species loss today is comparable to those periods of mass extinction. However, there is a major difference between the previous mass extinctions and the current extinction we are experiencing: human activity. Specifically, three human activities have a major impact: destruction of habitat, introduction of exotic species, and over-harvesting. Predictions of species loss within the next century, a tiny amount of time on geological timescales, range from 10 percent to 50 percent. Extinctions on this scale have only happened five other times in the history of the planet, and they have been caused by cataclysmic events that changed the course of the history of life in each instance. Earth is now in one of those times.

## Types of Biodiversity

Scientists generally accept that the term biodiversity describes the number and kinds of species in a location or on the planet. Species can be difficult to define, but most biologists still feel comfortable with the concept and are able to identify and count eukaryotic species in most contexts. Biologists have also identified alternate measures of biodiversity, some of which are important for planning how to preserve biodiversity.

**Genetic diversity** is one of those alternate concepts. Genetic diversity or variation is the raw material for adaptation in a species. A species' future potential for adaptation depends on the genetic diversity held in the

genomes of the individuals in populations that make up the species. The same is true for higher taxonomic categories. A genus with very different types of species will have more genetic diversity than a genus with species that look alike and have similar ecologies. If there were a choice between one of these genera of species being preserved, the one with the greatest potential for subsequent evolution is the most genetically diverse one. It would be ideal not to have to make such choices, but increasingly this may be the norm.

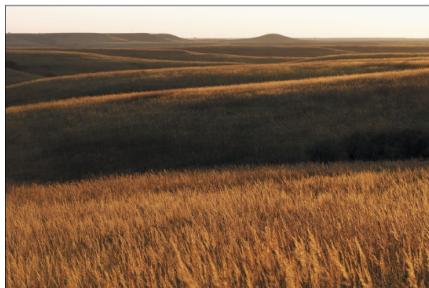
Many genes code for proteins, which in turn carry out the metabolic processes that keep organisms alive and reproducing. Genetic diversity can be measured as **chemical diversity** in that different species produce a variety of chemicals in their cells, both the proteins as well as the products and byproducts of metabolism. This chemical diversity has potential benefit for humans as a source of pharmaceuticals, so it provides one way to measure diversity that is important to human health and welfare.

Humans have generated diversity in domestic animals, plants, and fungi. This diversity is also suffering losses because of migration, market forces, and increasing globalism in agriculture, especially in heavily populated regions such as China, India, and Japan. The human population directly depends on this diversity as a stable food source, and its decline is troubling biologists and agricultural scientists.

It is also useful to define **ecosystem diversity**, meaning the number of different ecosystems on the planet or in a given geographic area ([link](#)). Whole ecosystems can disappear even if some of the species might survive by adapting to other ecosystems. The loss of an ecosystem means the loss of interactions between species, the loss of unique features of coadaptation, and the loss of biological productivity that an ecosystem is able to create. An example of a largely extinct ecosystem in North America is the prairie ecosystem. Prairies once spanned central North America from the boreal forest in northern Canada down into Mexico. They are now all but gone, replaced by crop fields, pasture lands, and suburban sprawl. Many of the species survive, but the hugely productive ecosystem that was responsible for creating the most productive agricultural soils is now gone. As a consequence, soils are disappearing or must be maintained at greater expense.



(a)



(b)

The variety of ecosystems on Earth—from (a) coral reef to (b) prairie—enables a great diversity of species to exist. (credit a: modification of work by Jim

Maragos, USFWS; credit b:  
modification of work by Jim  
Minnerath, USFWS)

## Current Species Diversity

Despite considerable effort, knowledge of the species that inhabit the planet is limited. A recent estimate suggests that the eukaryote species for which science has names, about 1.5 million species, account for less than 20 percent of the total number of eukaryote species present on the planet (8.7 million species, by one estimate). Estimates of numbers of prokaryotic species are largely guesses, but biologists agree that science has only begun to catalog their diversity. Even with what is known, there is no central repository of names or samples of the described species; therefore, there is no way to be sure that the 1.5 million descriptions is an accurate number. It is a best guess based on the opinions of experts in different taxonomic groups. Given that Earth is losing species at an accelerating pace, science is very much in the place it was with the Lake Victoria cichlids: knowing little about what is being lost. [\[link\]](#) presents recent estimates of biodiversity in different groups.

Estimates of the Numbers of Described and Predicted Species by Taxonomic Group				
	Mora et al. 2011 <sup>[footnote]</sup> Mora Camilo et al., “How Many Species Are There on Earth and in the Ocean?” <i>PLoS Biology</i> (2011), doi:10.1371/journal.pbio.1001127.		Chapman 2009 <sup>[footnote]</sup> Arthur D. Chapman, <i>Numbers of Living Species in</i> . 2nd ed. (Canberra, AU: Australian Biological Reso <a href="http://www.environment.gov.au/biodiversity/abrs/pnumbers/2009/pubs/nlsaw-2nd-complete.pdf">http://www.environment.gov.au/biodiversity/abrs/pnumbers/2009/pubs/nlsaw-2nd-complete.pdf</a> .	
	Described	Predicted	Described	Predicted
Animalia	1,124,516	9,920,000	1,424,153	6,836,330
Chromista	17,892	34,900	25,044	200,500
Fungi	44,368	616,320	98,998	1,500,000
Plantae	224,244	314,600	310,129	390,800
Protozoa	16,236	72,800	28,871	1,000,000
Prokaryotes	—	—	10,307	1,000,000
Total	1,438,769	10,960,000	1,897,502	10,897,630

There are various initiatives to catalog described species in accessible ways, and the internet is facilitating that effort. Nevertheless, it has been pointed out that at the current rate of species description, which according to the

State of Observed Species Report is 17,000 to 20,000 new species per year, it will take close to 500 years to finish describing life on this planet. [\[footnote\]](#) Over time, the task becomes both increasingly impossible and increasingly easier as extinction removes species from the planet.

International Institute for Species Exploration (IISE), *2011 State of Observed Species (SOS)*. Tempe, AZ: IISE, 2011. Accessed May, 20, 2012. <http://species.asu.edu/SOS>.

Naming and counting species may seem an unimportant pursuit given the other needs of humanity, but it is not simply an accounting. Describing species is a complex process by which biologists determine an organism's unique characteristics and whether or not that organism belongs to any other described species. It allows biologists to find and recognize the species after the initial discovery, and allows them to follow up on questions about its biology. In addition, the unique characteristics of each species make it potentially valuable to humans or other species on which humans depend. Understanding these characteristics is the value of finding and naming species.

## Patterns of Biodiversity

Biodiversity is not evenly distributed on Earth. Lake Victoria contained almost 500 species of cichlids alone, ignoring the other fish families present in the lake. All of these species were found only in Lake Victoria; therefore, the 500 species of cichlids were endemic. **Endemic species** are found in only one location. Endemics with highly restricted distributions are particularly vulnerable to extinction. Higher taxonomic levels, such as genera and families, can also be endemic. Lake Huron contains about 79 species of fish, all of which are found in many other lakes in North America. What accounts for the difference in fish diversity in these two lakes? Lake Victoria is a tropical lake, while Lake Huron is a temperate lake. Lake Huron in its present form is only about 7,000 years old, while Lake Victoria in its present form is about 15,000 years old. Biogeographers have suggested these two factors, latitude and age, are two of several hypotheses to explain biodiversity patterns on the planet.

### Note:

#### Career Connection

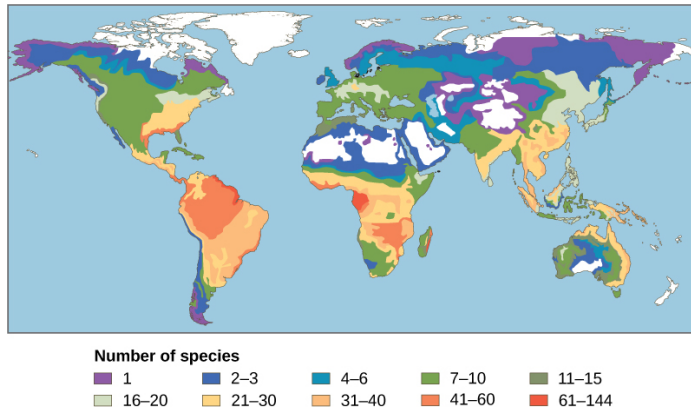
#### Biogeographer

Biogeography is the study of the distribution of the world's species—both in the past and in the present. The work of biogeographers is critical to understanding our physical environment, how the environment affects species, and how environmental changes impact the distribution of a species; it has also been critical to developing evolutionary theory. Biogeographers need to understand both biology and ecology. They also need to be well-versed in evolutionary studies, soil science, and climatology.

There are three main fields of study under the heading of biogeography: ecological biogeography, historical biogeography (called paleobiogeography), and conservation biogeography. Ecological biogeography studies the current factors affecting the distribution of plants and animals. Historical biogeography, as the name implies, studies the past distribution of species. Conservation biogeography, on the other hand, is focused on the protection and restoration of species based upon known historical and current ecological information. Each of these fields considers both zoogeography and phytogeography—the past and present distribution of animals and plants.

One of the oldest observed patterns in ecology is that species biodiversity in almost every taxonomic group increases as latitude declines. In other words, biodiversity increases closer to the equator ([\[link\]](#)).





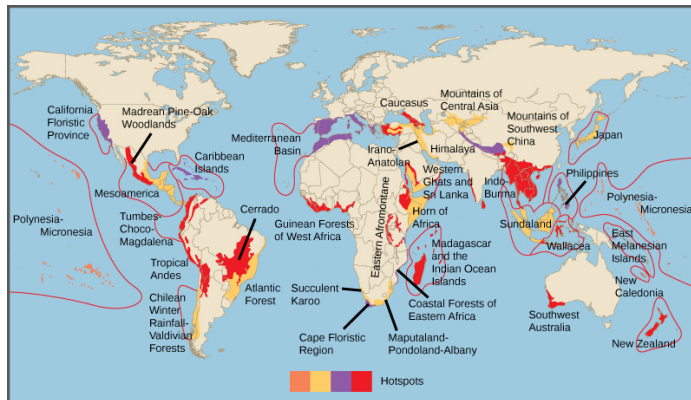
This map illustrates the number of amphibian species across the globe and shows the trend toward higher biodiversity at lower latitudes. A similar pattern is observed for most taxonomic groups.

It is not yet clear why biodiversity increases closer to the equator, but hypotheses include the greater age of the ecosystems in the tropics versus temperate regions that were largely devoid of life or drastically impoverished during the last glaciation. The idea is that greater age provides more time for speciation. Another possible explanation is the increased energy the tropics receive from the sun versus the decreased energy that temperate and polar regions receive. It is not entirely clear how greater energy input could translate into more species. The complexity of tropical ecosystems may promote speciation by increasing the **heterogeneity**, or number of ecological niches, in the tropics relative to higher latitudes. The greater heterogeneity provides more opportunities for coevolution, specialization, and perhaps greater selection pressures leading to population differentiation. However, this hypothesis suffers from some circularity—ecosystems with more species encourage speciation, but how did they get more species to begin with? The tropics have been perceived as being more stable than temperate regions, which have a pronounced climate and day-length seasonality. The tropics have their own forms of seasonality, such as rainfall, but they are generally assumed to be more stable environments and this stability might promote speciation.

Regardless of the mechanisms, it is certainly true that all levels of biodiversity are greatest in the tropics. Additionally, the rate of endemism is highest, and there are more biodiversity hotspots. However, this richness of diversity also means that knowledge of species is lowest, and there is a high potential for biodiversity loss.

### Conservation of Biodiversity

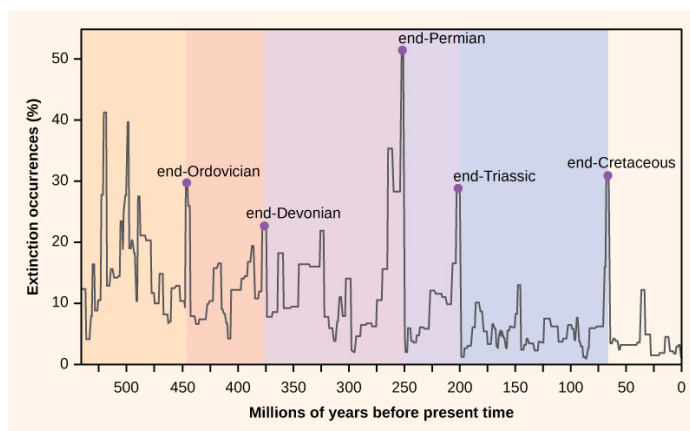
In 1988, British environmentalist Norman Myers developed a conservation concept to identify areas rich in species and at significant risk for species loss: biodiversity hotspots. **Biodiversity hotspots** are geographical areas that contain high numbers of endemic species. The purpose of the concept was to identify important locations on the planet for conservation efforts, a kind of conservation triage. By protecting hotspots, governments are able to protect a larger number of species. The original criteria for a hotspot included the presence of 1500 or more endemic plant species and 70 percent of the area disturbed by human activity. There are now 34 biodiversity hotspots ([link](#)) containing large numbers of endemic species, which include half of Earth's endemic plants.



Conservation International has identified 34 biodiversity hotspots, which cover only 2.3 percent of the Earth's surface but have endemic to them 42 percent of the terrestrial vertebrate species and 50 percent of the world's plants.

## Biodiversity Change through Geological Time

The number of species on the planet, or in any geographical area, is the result of an equilibrium of two evolutionary processes that are ongoing: speciation and extinction. Both are natural “birth” and “death” processes of macroevolution. When speciation rates begin to outstrip extinction rates, the number of species will increase; likewise, the number of species will decrease when extinction rates begin to overtake speciation rates. Throughout Earth's history, these two processes have fluctuated—sometimes leading to dramatic changes in the number of species on Earth as reflected in the fossil record ([link](#)).



Percent extinction occurrences as reflected in the fossil record have fluctuated throughout Earth's history. Sudden and dramatic losses of biodiversity, called mass extinctions, have occurred five times.



Paleontologists have identified five strata in the fossil record that appear to show sudden and dramatic (greater than half of all extant species disappearing from the fossil record) losses in biodiversity. These are called mass extinctions. There are many lesser, yet still dramatic, extinction events, but the five mass extinctions have attracted the most research. An argument can be made that the five mass extinctions are only the five most extreme events in a continuous series of large extinction events throughout the Phanerozoic (since 542 million years ago). In most cases, the hypothesized causes are still controversial; however, the most recent event seems clear.

**The Five Mass Extinctions**

The fossil record of the mass extinctions was the basis for defining periods of geological history, so they typically occur at the transition point between geological periods. The transition in fossils from one period to another reflects the dramatic loss of species and the gradual origin of new species. These transitions can be seen in the rock strata. [\[link\]](#) provides data on the five mass extinctions.

Mass Extinctions		
Geological Period	Mass Extinction Name	Time (millions of years ago)
Ordovician–Silurian	end-Ordovician O–S	450–440
Late Devonian	end-Devonian	375–360
Permian–Triassic	end-Permian	251
Triassic–Jurassic	end-Triassic	205
Cretaceous–Paleogene	end-Cretaceous K–Pg (K–T)	65.5

This table shows the names and dates for the five mass extinctions in Earth’s history.

The Ordovician-Silurian extinction event is the first recorded mass extinction and the second largest. During this period, about 85 percent of marine species (few species lived outside the oceans) became extinct. The main hypothesis for its cause is a period of glaciation and then warming. The extinction event actually consists of two extinction events separated by about 1 million years. The first event was caused by cooling, and the second event was due to the subsequent warming. The climate changes affected temperatures and sea levels. Some researchers have suggested that a gamma-ray burst, caused by a nearby supernova, is a possible cause of the Ordovician-Silurian extinction. The gamma-ray burst would have stripped away the Earth’s ozone layer causing intense ultraviolet radiation from the sun and may account for climate changes observed at the time. The hypothesis is speculative, but extraterrestrial influences on Earth’s history are an active line of research. Recovery of biodiversity after the mass extinction took from 5 to 20 million years, depending on the location.

The late Devonian extinction may have occurred over a relatively long period of time. It appears to have affected marine species and not the plants or animals inhabiting terrestrial habitats. The causes of this extinction are poorly understood.

The end-Permian extinction was the largest in the history of life. Indeed, an argument could be made that Earth nearly became devoid of life during this extinction event. The planet looked very different before and after this event. Estimates are that 96 percent of all marine species and 70 percent of all terrestrial species were lost. It was at this time, for example, that the trilobites, a group that survived the Ordovician–Silurian extinction, became extinct. The causes for this mass extinction are not clear, but the leading suspect is extended and widespread

volcanic activity that led to a runaway global-warming event. The oceans became largely anoxic, suffocating marine life. Terrestrial tetrapod diversity took 30 million years to recover after the end-Permian extinction. The Permian extinction dramatically altered Earth's biodiversity makeup and the course of evolution.

The causes of the Triassic–Jurassic extinction event are not clear and hypotheses of climate change, asteroid impact, and volcanic eruptions have been argued. The extinction event occurred just before the breakup of the supercontinent Pangaea, although recent scholarship suggests that the extinctions may have occurred more gradually throughout the Triassic.

The causes of the end-Cretaceous extinction event are the ones that are best understood. It was during this extinction event about 65 million years ago that the dinosaurs, the dominant vertebrate group for millions of years, disappeared from the planet (with the exception of a theropod clade that gave rise to birds). Indeed, every land animal that weighed more than 25 kg became extinct. The cause of this extinction is now understood to be the result of a cataclysmic impact of a large meteorite, or asteroid, off the coast of what is now the Yucatán Peninsula. This hypothesis, proposed first in 1980, was a radical explanation based on a sharp spike in the levels of iridium (which rains down from space in meteors at a fairly constant rate but is otherwise absent on Earth's surface) at the rock stratum that marks the boundary between the Cretaceous and Paleogene periods ([link](#)). This boundary marked the disappearance of the dinosaurs in fossils as well as many other taxa. The researchers who discovered the iridium spike interpreted it as a rapid influx of iridium from space to the atmosphere (in the form of a large asteroid) rather than a slowing in the deposition of sediments during that period. It was a radical explanation, but the report of an appropriately aged and sized impact crater in 1991 made the hypothesis more believable. Now an abundance of geological evidence supports the theory. Recovery times for biodiversity after the end-Cretaceous extinction are shorter, in geological time, than for the end-Permian extinction, on the order of 10 million years.

**Note:**

**Art Connection**



In 1980, Luis and Walter Alvarez, Frank Asaro, and Helen Michels discovered, across the world, a spike in the concentration of iridium within the sedimentary layer at the K–Pg boundary. These researchers hypothesized that this iridium spike was caused by an asteroid impact that resulted in the K–Pg mass extinction. In the photo, the iridium layer is the light band. (credit: USGS)

Scientists measured the relative abundance of fern spores above and below the K–Pg boundary in this rock sample. Which of the following statements most likely represents their findings?

- a. An abundance of fern spores from several species was found below the K–Pg boundary, but none was found above.
- b. An abundance of fern spores from several species was found above the K–Pg boundary, but none was found below.
- c. An abundance of fern spores was found both above and below the K–Pg boundary, but only one species was found below the boundary, and many species were found above the boundary.
- d. Many species of fern spores were found both above and below the boundary, but the total number of spores was greater below the boundary.

**Note:**

**Link to Learning**



Explore this [interactive website](#) about mass extinctions.

## The Pleistocene Extinction

The Pleistocene Extinction is one of the lesser extinctions, and a recent one. It is well known that the North American, and to some degree Eurasian, **megafauna**, or large animals, disappeared toward the end of the last glaciation period. The extinction appears to have happened in a relatively restricted time period of 10,000–12,000 years ago. In North America, the losses were quite dramatic and included the woolly mammoths (last dated about 4,000 years ago in an isolated population), mastodon, giant beavers, giant ground sloths, saber-toothed cats, and the North American camel, just to name a few. The possibility that the rapid extinction of these large animals was caused by over-hunting was first suggested in the 1900s. Research into this hypothesis continues today. It seems likely that over-hunting caused many pre-written history extinctions in many regions of the world.

In general, the timing of the Pleistocene extinctions correlated with the arrival of humans and not with climate-change events, which is the main competing hypothesis for these extinctions. The extinctions began in Australia about 40,000 to 50,000 years ago, just after the arrival of humans in the area: a marsupial lion, a giant one-ton wombat, and several giant kangaroo species disappeared. In North America, the extinctions of almost all of the large mammals occurred 10,000–12,000 years ago. All that are left are the smaller mammals such as bears, elk, moose, and cougars. Finally, on many remote oceanic islands, the extinctions of many species occurred coincident with human arrivals. Not all of the islands had large animals, but when there were large animals, they were lost. Madagascar was colonized about 2,000 years ago and the large mammals that lived there became extinct. Eurasia and Africa do not show this pattern, but they also did not experience a recent arrival of humans. Humans arrived in Eurasia hundreds of thousands of years ago after the origin of the species in Africa. This topic remains an area of active research and hypothesizing. It seems clear that even if climate played a role, in most cases human hunting precipitated the extinctions.

## Present-Time Extinctions

The sixth, or Holocene, mass extinction appears to have begun earlier than previously believed and has mostly to do with the activities of *Homo sapiens*. Since the beginning of the Holocene period, there are numerous recent extinctions of individual species that are recorded in human writings. Most of these are coincident with the expansion of the European colonies since the 1500s.

One of the earlier and popularly known examples is the dodo bird. The dodo bird lived in the forests of Mauritius, an island in the Indian Ocean. The dodo bird became extinct around 1662. It was hunted for its meat by sailors and was easy prey because the dodo, which did not evolve with humans, would approach people without fear. Introduced pigs, rats, and dogs brought to the island by European ships also killed dodo young and eggs.

Steller's sea cow became extinct in 1768; it was related to the manatee and probably once lived along the northwest coast of North America. Steller's sea cow was first discovered by Europeans in 1741 and was hunted for meat and oil. The last sea cow was killed in 1768. That amounts to 27 years between the sea cow's first contact with Europeans and extinction of the species.

In 1914, the last living passenger pigeon died in a zoo in Cincinnati, Ohio. This species had once darkened the skies of North America during its migrations, but it was hunted and suffered from habitat loss through the clearing of forests for farmland. In 1918, the last living Carolina parakeet died in captivity. This species was once common in the eastern United States, but it suffered from habitat loss. The species was also hunted because it ate orchard fruit when its native foods were destroyed to make way for farmland. The Japanese sea lion, which inhabited a broad area around Japan and the coast of Korea, became extinct in the 1950s due to fishermen. The Caribbean monk seal was distributed throughout the Caribbean Sea but was driven to extinction via hunting by 1952.

These are only a few of the recorded extinctions in the past 500 years. The International Union for Conservation of Nature (IUCN) keeps a list of extinct and endangered species called the Red List. The list is not complete, but it describes 380 extinct species of vertebrates after 1500 AD, 86 of which were driven extinct by overhunting or overfishing.

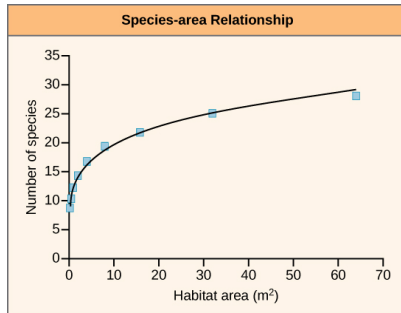
### Estimates of Present-Time Extinction Rates

Estimates of **extinction rates** are hampered by the fact that most extinctions are probably happening without observation. The extinction of a bird or mammal is likely to be noticed by humans, especially if it has been hunted or used in some other way. But there are many organisms that are of less interest to humans (not necessarily of less value) and many that are undescribed.

The background extinction rate is estimated to be about one per million species per year (E/MSY). For example, assuming there are about ten million species in existence, the expectation is that ten species would become extinct each year (each year represents ten million species per year).

One contemporary extinction rate estimate uses the extinctions in the written record since the year 1500. For birds alone this method yields an estimate of 26 E/MSY. However, this value may be underestimated for three reasons. First, many species would not have been described until much later in the time period, so their loss would have gone unnoticed. Second, the number of recently extinct species is increasing because extinct species now are being described from skeletal remains. And third, some species are probably already extinct even though conservationists are reluctant to name them as such. Taking these factors into account raises the estimated extinction rate closer to 100 E/MSY. The predicted rate by the end of the century is 1500 E/MSY.

A second approach to estimating present-time extinction rates is to correlate species loss with habitat loss by measuring forest-area loss and understanding species-area relationships. The **species-area relationship** is the rate at which new species are seen when the area surveyed is increased. Studies have shown that the number of species present increases as the size of the island increases. This phenomenon has also been shown to hold true in other habitats as well. Turning this relationship around, if the habitat area is reduced, the number of species living there will also decline. Estimates of extinction rates based on habitat loss and species-area relationships have suggested that with about 90 percent habitat loss an expected 50 percent of species would become extinct. Species-area estimates have led to species extinction rate calculations of about 1000 E/MSY and higher. In general, actual observations do not show this amount of loss and suggestions have been made that there is a delay in extinction. Recent work has also called into question the applicability of the species-area relationship when estimating the loss of species. This work argues that the species-area relationship leads to an overestimate of extinction rates. A better relationship to use may be the endemics-area relationship. Using this method would bring estimates down to around 500 E/MSY in the coming century. Note that this value is still 500 times the background rate.



Studies have shown that the number of species present increases with the size of the habitat. (credit: modification of work by Adam B. Smith)

#### Note:

Link to Learning



Check out this [interactive exploration](#) of endangered and extinct species, their ecosystems, and the causes of the endangerment or extinction.

## Section Summary

Biodiversity exists at multiple levels of organization and is measured in different ways depending on the goals of those taking the measurements. These measurements include numbers of species, genetic diversity, chemical diversity, and ecosystem diversity. The number of described species is estimated to be 1.5 million with about 17,000 new species being described each year. Estimates for the total number of species on Earth vary but are on the order of 10 million. Biodiversity is negatively correlated with latitude for most taxa, meaning that biodiversity is higher in the tropics. The mechanism for this pattern is not known with certainty, but several plausible hypotheses have been advanced.

Five mass extinctions with losses of more than 50 percent of extant species are observable in the fossil record. Biodiversity recovery times after mass extinctions vary, but have been up to 30 million years. Recent extinctions are recorded in written history and are the basis for one method of estimating contemporary extinction rates. The other method uses measures of habitat loss and species-area relationships. Estimates of contemporary extinction rates vary, but some rates are as high as 500 times the background rate, as determined from the fossil record, and are predicted to rise.

## Art Connections

### Exercise:

**Problem:**

[\[link\]](#) Scientists measured the relative abundance of fern spores above and below the K-Pg boundary in this rock sample. Which of the following statements most likely represents their findings?

- a. An abundance of fern spores from several species was found below the K-Pg boundary, but none was found above.
- b. An abundance of fern spores from several species was found above the K-Pg boundary, but none was found below.
- c. An abundance of fern spores was found both above and below the K-Pg boundary, but only one species was found below the boundary, and many species were found above the boundary.
- d. Many species of fern spores were found both above and below the boundary, but the total number of spores was greater below the boundary.

---

**Solution:**

[\[link\]](#) A. An abundance of fern spores from several species was found below the K-Pg boundary, but none was found above.

**Review Questions**

**Exercise:**

**Problem:**

With an extinction rate of 100 E/MSY and an estimated 10 million species, how many extinctions are expected to occur in a century?

- a. 100
- b. 10, 000
- c. 100, 000
- d. 1, 000, 000

---

**Solution:**

C

**Exercise:**

**Problem:** An adaptive radiation is\_\_\_\_\_.

- a. a burst of speciation
- b. a healthy level of UV radiation
- c. a hypothesized cause of a mass extinction
- d. evidence of an asteroid impact

---

**Solution:**

A

**Exercise:**

**Problem:** The number of currently described species on the planet is about \_\_\_\_\_.

- a. 17,000
- b. 150,000
- c. 1.5 million
- d. 10 million

---

**Solution:**

C

**Exercise:**

**Problem:** A mass extinction is defined as \_\_\_\_\_.

- a. a loss of 95 percent of species
- b. an asteroid impact
- c. a boundary between geological periods
- d. a loss of 50 percent of species

---

**Solution:**

D

## Free Response

**Exercise:**

**Problem:** Describe the evidence for the cause of the Cretaceous–Paleogene (K–Pg) mass extinction.

---

**Solution:**

The hypothesized cause of the K–Pg extinction event is an asteroid impact. The first piece of evidence of the impact is a spike in iridium (an element that is rare on Earth, but common in meteors) in the geological layers that mark the K–Pg transition. The second piece of evidence is an impact crater off the Yucatán Peninsula that is the right size and age to have caused the extinction event.

**Exercise:**

**Problem:** Describe the two methods used to calculate contemporary extinction rates.

---

**Solution:**

Extinction rates are calculated based on the recorded extinction of species in the past 500 years. Adjustments are made for unobserved extinctions and undiscovered species. The second method is a calculation based on the amount of habitat destruction and species-area curves.

## Glossary

adaptive radiation

rapid branching through speciation of a phylogenetic tree into many closely related species

biodiversity

variety of a biological system, typically conceived as the number of species, but also applying to genes, biochemistry, and ecosystems

biodiversity hotspot

concept originated by Norman Myers to describe a geographical region with a large number of endemic species and a large percentage of degraded habitat

chemical diversity

variety of metabolic compounds in an ecosystem

ecosystem diversity

variety of ecosystems

endemic species

species native to one place

extinction

disappearance of a species from Earth; local extinction is the disappearance of a species from a region

extinction rate

number of species becoming extinct over time, sometimes defined as extinctions per million species–years to make numbers manageable (E/MSY)

genetic diversity

variety of genes in a species or other taxonomic group or ecosystem, the term can refer to allelic diversity or genome-wide diversity

heterogeneity

number of ecological niches

megafauna

large animals

species-area relationship

relationship between area surveyed and number of species encountered; typically measured by incrementally increasing the area of a survey and determining the cumulative numbers of species



## The Importance of Biodiversity to Human Life

By the end of this section, you will be able to:

- Identify chemical diversity benefits to humans
- Identify biodiversity components that support human agriculture
- Describe ecosystem services

It may not be clear why biologists are concerned about biodiversity loss. When biodiversity loss is thought of as the extinction of the passenger pigeon, the dodo bird, and even the woolly mammoth, the loss may appear to be an emotional one. But is the loss practically important for the welfare of the human species? From the perspective of evolution and ecology, the loss of a particular individual species is unimportant (however, the loss of a keystone species can lead to ecological disaster). Extinction is a normal part of macroevolution. But the accelerated extinction rate means the loss of tens of thousands of species within our lifetimes, and it is likely to have dramatic effects on human welfare through the collapse of ecosystems and in added costs to maintain food production, clean air and water, and human health.

Agriculture began after early hunter-gatherer societies first settled in one place and heavily modified their immediate environment. This cultural transition has made it difficult for humans to recognize their dependence on undomesticated living things on the planet. Biologists recognize the human species is embedded in ecosystems and is dependent on them, just as every other species on the planet is dependent. Technology smoothes out the extremes of existence, but ultimately the human species cannot exist without its ecosystem.

## Human Health

Contemporary societies that live close to the land often have a broad knowledge of the medicinal uses of plants growing in their area. Most plants produce **secondary plant compounds**, which are toxins used to protect the plant from insects and other animals that eat them, but some of which also work as medication. For centuries in Europe, older knowledge about the medical uses of plants was compiled in herbals—books that

identified plants and their uses. Humans are not the only species to use plants for medicinal reasons: the great apes, orangutans, chimpanzees, bonobos, and gorillas have all been observed self-medicating with plants.

Modern pharmaceutical science also recognizes the importance of these plant compounds. Examples of significant medicines derived from plant compounds include aspirin, codeine, digoxin, atropine, and vincristine ([link](#)). Many medicines were once derived from plant extracts but are now synthesized. It is estimated that, at one time, 25 percent of modern drugs contained at least one plant extract. That number has probably decreased to about 10 percent as natural plant ingredients are replaced by synthetic versions. Antibiotics, which are responsible for extraordinary improvements in health and lifespans in developed countries, are compounds largely derived from fungi and bacteria.



*Catharanthus roseus*, the Madagascar periwinkle, has various medicinal properties. Among other uses, it is a source of vincristine, a drug used in the treatment of lymphomas.  
(credit: Forest and Kim Starr)

In recent years, animal venoms and poisons have excited intense research for their medicinal potential. By 2007, the FDA had approved five drugs based on animal toxins to treat diseases such as hypertension, chronic pain, and diabetes. Another five drugs are undergoing clinical trials, and at least six drugs are being used in other countries. Other toxins under investigation come from mammals, snakes, lizards, various amphibians, fish, snails, octopuses, and scorpions.

Aside from representing billions of dollars in profits, these medicines improve people's lives. Pharmaceutical companies are actively looking for new compounds synthesized by living organisms that can function as medicine. It is estimated that 1/3 of pharmaceutical research and development is spent on natural compounds and that about 35 percent of new drugs brought to market between 1981 and 2002 were from natural compounds. The opportunities for new medications will be reduced in direct proportion to the disappearance of species.

## **Agricultural Diversity**

Since the beginning of human agriculture more than 10,000 years ago, human groups have been breeding and selecting crop varieties. This crop diversity matched the cultural diversity of highly subdivided populations of humans. For example, potatoes were domesticated beginning around 7,000 years ago in the central Andes of Peru and Bolivia. The potatoes grown in that region belong to seven species and the number of varieties likely is in the thousands. Each variety has been bred to thrive at particular elevations and soil and climate conditions. The diversity is driven by the diverse demands of the topography, the limited movement of people, and the demands created by crop rotation for different varieties that will do well in different fields.

Potatoes are only one example of human-generated diversity. Every plant, animal, and fungus that has been cultivated by humans has been bred from original wild ancestor species into diverse varieties arising from the demands for food value, adaptation to growing conditions, and resistance to pests. The potato demonstrates a well-known example of the risks of low crop diversity: the tragic Irish potato famine when the single variety grown

in Ireland became susceptible to a potato blight, wiping out the crop. The loss of the crop led to famine, death, and mass emigration. Resistance to disease is a chief benefit to maintaining crop biodiversity, and lack of diversity in contemporary crop species carries similar risks. Seed companies, which are the source of most crop varieties in developed countries, must continually breed new varieties to keep up with evolving pest organisms. These same seed companies, however, have participated in the decline of the number of varieties available as they focus on selling fewer varieties in more areas of the world.

The ability to create new crop varieties relies on the diversity of varieties available and the accessibility of wild forms related to the crop plant. These wild forms are often the source of new gene variants that can be bred with existing varieties to create varieties with new attributes. Loss of wild species related to a crop will mean the loss of potential in crop improvement. Maintaining the genetic diversity of wild species related to domesticated species ensures our continued food supply.

Since the 1920s, government agriculture departments have maintained seed banks of crop varieties as a way to maintain crop diversity. This system has flaws because, over time, seed banks are lost through accidents, and there is no way to replace them. In 2008, the Svalbard Global Seed Vault ([\[link\]](#)) began storing seeds from around the world as a backup system to the regional seed banks. If a regional seed bank stores varieties in Svalbard, losses can be replaced from Svalbard. The seed vault is located deep into the rock of an arctic island. Conditions within the vault are maintained at ideal temperature and humidity for seed survival, but the deep underground location of the vault in the arctic means that failure of the vault's systems will not compromise the climatic conditions inside the vault.

**Note:**

Art Connection



The Svalbard Global Seed Vault is a storage facility for seeds of Earth's diverse crops. (credit: Mari Tefre, Svalbard Global Seed Vault)

The Svalbard Global Seed Vault is located on Spitsbergen island in Norway, which has an arctic climate. Why might an arctic climate be good for seed storage?

Crop success is largely dependent on the quality of the soil. Although some agricultural soils are rendered sterile using controversial cultivation and chemical treatments, most contain a huge diversity of organisms that maintain nutrient cycles—breaking down organic matter into nutrient compounds that crops need for growth. These organisms also maintain soil texture that affects water and oxygen dynamics in the soil that are necessary for plant growth. If farmers had to maintain arable soil using alternate means, the cost of food would be much higher than it is now. These kinds of processes are called ecosystem services. They occur within ecosystems, such as soil ecosystems, as a result of the diverse metabolic activities of the organisms living there, but they provide benefits to human food production, drinking water availability, and breathable air.

Other key ecosystem services related to food production are plant pollination and crop pest control. Over 150 crops in the United States require pollination to produce. One estimate of the benefit of honeybee pollination within the United States is \$1.6 billion per year; other pollinators contribute up to \$6.7 billion more.

Many honeybee populations are managed by apiarists who rent out their hives' services to farmers. Honeybee populations in North America have been suffering large losses caused by a syndrome known as colony collapse disorder, whose cause is unclear. Other pollinators include a diverse array of other bee species and various insects and birds. Loss of these species would make growing crops requiring pollination impossible, increasing dependence on other crops.

Finally, humans compete for their food with crop pests, most of which are insects. Pesticides control these competitors; however, pesticides are costly and lose their effectiveness over time as pest populations adapt. They also lead to collateral damage by killing non-pest species and risking the health of consumers and agricultural workers. Ecologists believe that the bulk of the work in removing pests is actually done by predators and parasites of those pests, but the impact has not been well studied. A review found that in 74 percent of studies that looked for an effect of landscape complexity on natural enemies of pests, the greater the complexity, the greater the effect of pest-suppressing organisms. An experimental study found that introducing multiple enemies of pea aphids (an important alfalfa pest) increased the yield of alfalfa significantly. This study shows the importance of landscape diversity via the question of whether a diversity of pests is more effective at control than one single pest; the results showed this to be the case. Loss of diversity in pest enemies will inevitably make it more difficult and costly to grow food.

## **Wild Food Sources**

In addition to growing crops and raising animals for food, humans obtain food resources from wild populations, primarily fish populations. For approximately 1 billion people, aquatic resources provide the main source of animal protein. But since 1990, global fish production has declined.

Despite considerable effort, few fisheries on the planet are managed for sustainability.

Fishery extinctions rarely lead to complete extinction of the harvested species, but rather to a radical restructuring of the marine ecosystem in which a dominant species is so over-harvested that it becomes a minor player, ecologically. In addition to humans losing the food source, these alterations affect many other species in ways that are difficult or impossible to predict. The collapse of fisheries has dramatic and long-lasting effects on local populations that work in the fishery. In addition, the loss of an inexpensive protein source to populations that cannot afford to replace it will increase the cost of living and limit societies in other ways. In general, the fish taken from fisheries have shifted to smaller species as larger species are fished to extinction. The ultimate outcome could clearly be the loss of aquatic systems as food sources.

**Note:**

Link to Learning



View a [brief video](#) discussing declining fish stocks.

## Psychological and Moral Value

Finally, it has been argued that humans benefit psychologically from living in a biodiverse world. A chief proponent of this idea is entomologist E. O. Wilson. He argues that human evolutionary history has adapted us to live in a natural environment and that built environments generate stressors that affect human health and well-being. There is considerable research into the

psychological regenerative benefits of natural landscapes that suggests the hypothesis may hold some truth. In addition, there is a moral argument that humans have a responsibility to inflict as little harm as possible on other species.

## Section Summary

Humans use many compounds that were first discovered or derived from living organisms as medicines: secondary plant compounds, animal toxins, and antibiotics produced by bacteria and fungi. More medicines are expected to be discovered in nature. Loss of biodiversity will impact the number of pharmaceuticals available to humans.

Crop diversity is a requirement for food security, and it is being lost. The loss of wild relatives to crops also threatens breeders' abilities to create new varieties. Ecosystems provide ecosystem services that support human agriculture: pollination, nutrient cycling, pest control, and soil development and maintenance. Loss of biodiversity threatens these ecosystem services and risks making food production more expensive or impossible. Wild food sources are mainly aquatic, but few are being managed for sustainability. Fisheries' ability to provide protein to human populations is threatened when extinction occurs.

Biodiversity may provide important psychological benefits to humans. Additionally, there are moral arguments for the maintenance of biodiversity.

## Art Connections

### Exercise:

#### Problem:

[\[link\]](#) The Svalbard Global Seed Vault is located on Spitsbergen island in Norway, which has an arctic climate. Why might an arctic climate be good for seed storage?

---

#### Solution:



[\[link\]](#) The ground is permanently frozen so the seeds will keep even if the electricity fails.

## Review Questions

### Exercise:

#### Problem:

A secondary plant compound might be used for which of the following?

- a. a new crop variety
- b. a new drug
- c. a soil nutrient
- d. a pest of a crop pest

---

#### Solution:

B

### Exercise:

**Problem:** Pollination is an example of \_\_\_\_\_.

- a. a possible source of new drugs
- b. chemical diversity
- c. an ecosystem service
- d. crop pest control

---

#### Solution:

C

### Exercise:

**Problem:**

What is an ecosystem service that performs the same function as a pesticide?

- a. pollination
- b. secondary plant compounds
- c. crop diversity
- d. predators of pests

---

**Solution:**

D

**Free Response****Exercise:**

**Problem:** Explain how biodiversity loss can impact crop diversity.

---

**Solution:**

Crop plants are derived from wild plants, and genes from wild relatives are frequently brought into crop varieties by plant breeders to add valued characteristics to the crops. If the wild species are lost, then this genetic variation would no longer be available.

**Exercise:****Problem:**

Describe two types of compounds from living things that are used as medications.

---

**Solution:**

Secondary plant compounds are toxins produced by plants to kill predators trying to eat them; some of these compounds can be used as drugs. Animal toxins such as snake venom can also be used as drugs. (Alternate answer: antibiotics are compounds produced by bacteria and fungi which can be used to kill bacteria.)

## **Glossary**

secondary plant compound

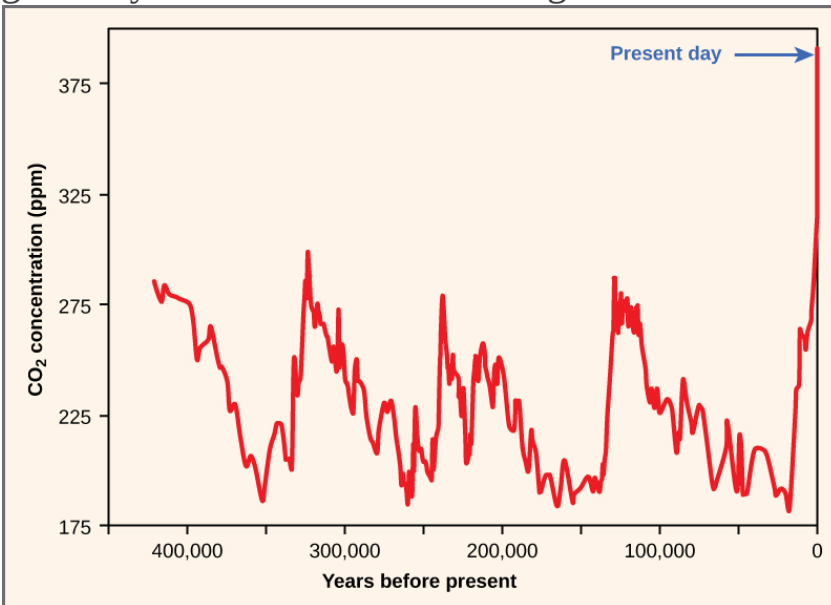
compound produced as byproducts of plant metabolic processes that is usually toxic, but is sequestered by the plant to defend against herbivores

## Threats to Biodiversity

By the end of this section, you will be able to:

- Identify significant threats to biodiversity
- Explain the effects of habitat loss, exotic species, and hunting on biodiversity
- Identify the early and predicted effects of climate change on biodiversity

The core threat to biodiversity on the planet, and therefore a threat to human welfare, is the combination of human population growth and resource exploitation. The human population requires resources to survive and grow, and those resources are being removed unsustainably from the environment. The three greatest proximate threats to biodiversity are habitat loss, overharvesting, and introduction of exotic species. The first two of these are a direct result of human population growth and resource use. The third results from increased mobility and trade. A fourth major cause of extinction, anthropogenic climate change, has not yet had a large impact, but it is predicted to become significant during this century. Global climate change is also a consequence of human population needs for energy and the use of fossil fuels to meet those needs ([\[link\]](#)). Environmental issues, such as toxic pollution, have specific targeted effects on species, but they are not generally seen as threats at the magnitude of the others.



Atmospheric carbon dioxide levels fluctuate in a cyclical manner. However, the burning of fossil fuels in recent history has caused a dramatic increase in the levels of carbon dioxide in the Earth's atmosphere, which have now reached levels never before seen in human history. Scientists predict that the addition of this “greenhouse gas” to the atmosphere is resulting in climate change that will significantly impact biodiversity in the coming century.

## **Habitat Loss**

Humans rely on technology to modify their environment and replace certain functions that were once performed by the natural ecosystem. Other species cannot do this. Elimination of their ecosystem—whether it is a forest, a desert, a grassland, a freshwater estuarine, or a marine environment—will kill the individuals in the species. Remove the entire habitat within the range of a species and, unless they are one of the few species that do well in human-built environments, the species will become extinct. Human destruction of habitats accelerated in the latter half of the twentieth century. Consider the exceptional biodiversity of Sumatra: it is home to one species of orangutan, a species of critically endangered elephant, and the Sumatran tiger, but half of Sumatra's forest is now gone. The neighboring island of Borneo, home to the other species of orangutan, has lost a similar area of forest. Forest loss continues in protected areas of Borneo. The orangutan in Borneo is listed as endangered by the International Union for Conservation of Nature (IUCN), but it is simply the most visible of thousands of species that will not survive the disappearance of the forests of Borneo. The forests are removed for timber and to plant palm oil plantations ([\[link\]](#)). Palm oil is used in many products including food products, cosmetics, and biodiesel in Europe. A five-year estimate of global forest cover loss for the years 2000–2005 was 3.1 percent. In the humid tropics where forest loss is primarily from timber extraction, 272,000 km<sup>2</sup> was lost out of a global total of

11,564,000 km<sup>2</sup> (or 2.4 percent). In the tropics, these losses certainly also represent the extinction of species because of high levels of endemism.



(a)



(b)



(c)



(d)



(e)

(a) One species of orangutan, *Pongo pygmaeus*, is found only in the rainforests of Borneo, and the other species of orangutan (*Pongo abelii*) is found only in the rainforests of Sumatra. These animals are examples of the exceptional biodiversity of

(c) the islands of Sumatra and Borneo. Other species include the (b) Sumatran tiger (*Panthera tigris sumatrae*) and the (d) Sumatran elephant (*Elephas maximus sumatranus*), both critically endangered species. Rainforest habitat is being removed to make way for (e) oil palm plantations such as this one in Borneo's Sabah Province. (credit a: modification of work by Thorsten Bachner; credit b: modification of work by Dick Mudde; credit c: modification of work by U.S. CIA World Factbook; credit d: modification of work by "Nonprofit Organizations"/Flickr; credit e: modification of work by Dr. Lian Pin Koh)

**Note:**

**Everyday Connection**

**Preventing Habitat Destruction with Wise Wood Choices**

Most consumers do not imagine that the home improvement products they buy might be contributing to habitat loss and species extinctions. Yet the market for illegally harvested tropical timber is huge, and the wood products often find themselves in building supply stores in the United States. One estimate is that 10 percent of the imported timber stream in the United States, which is the world's largest consumer of wood products, is potentially illegally logged. In 2006, this amounted to \$3.6 billion in wood products. Most of the illegal products are imported from countries that act as intermediaries and are not the originators of the wood.

How is it possible to determine if a wood product, such as flooring, was harvested sustainably or even legally? The Forest Stewardship Council (FSC) certifies sustainably harvested forest products, therefore, looking for their certification on flooring and other hardwood products is one way to ensure that the wood has not been taken illegally from a tropical forest.

Certification applies to specific products, not to a producer; some producers' products may not have certification while other products are certified. While there are other industry-backed certifications other than the FSC, these are unreliable due to lack of independence from the industry. Another approach is to buy domestic wood species. While it would be great if there was a list of legal versus illegal wood products, it is not that simple. Logging and forest management laws vary from country to country; what is illegal in one country may be legal in another. Where and how a product is harvested and whether the forest from which it comes is being maintained sustainably all factor into whether a wood product will be certified by the FSC. It is always a good idea to ask questions about where a wood product came from and how the supplier knows that it was harvested legally.

Habitat destruction can affect ecosystems other than forests. Rivers and streams are important ecosystems and are frequently modified through land development and from damming or water removal. Damming of rivers affects the water flow and access to all parts of a river. Differing flow regimes can reduce or eliminate populations that are adapted to these changes in flow patterns. For example, an estimated 91percent of river lengths in the United States have been developed: they have modifications like dams, to create energy or store water; levees, to prevent flooding; or dredging or rerouting, to create land that is more suitable for human development. Many fish species in the United States, especially rare species or species with restricted distributions, have seen declines caused by river damming and habitat loss. Research has confirmed that species of amphibians that must carry out parts of their life cycles in both aquatic and terrestrial habitats have a greater chance of suffering population declines and extinction because of the increased likelihood that one of their habitats or access between them will be lost.

## **Overharvesting**

Overharvesting is a serious threat to many species, but particularly to aquatic species. There are many examples of regulated commercial fisheries



monitored by fisheries scientists that have nevertheless collapsed. The western Atlantic cod fishery is the most spectacular recent collapse. While it was a hugely productive fishery for 400 years, the introduction of modern factory trawlers in the 1980s and the pressure on the fishery led to it becoming unsustainable. The causes of fishery collapse are both economic and political in nature. Most fisheries are managed as a common (shared) resource even when the fishing territory lies within a country's territorial waters. Common resources are subject to an economic pressure known as the **tragedy of the commons** in which essentially no fisher has a motivation to exercise restraint in harvesting a fishery when it is not owned by that fisher. The natural outcome of harvests of resources held in common is their overexploitation. While large fisheries are regulated to attempt to avoid this pressure, it still exists in the background. This overexploitation is exacerbated when access to the fishery is open and unregulated and when technology gives fishers the ability to overfish. In a few fisheries, the biological growth of the resource is less than the potential growth of the profits made from fishing if that time and money were invested elsewhere. In these cases—whales are an example—economic forces will always drive toward fishing the population to extinction.

**Note:**

Link to Learning



Explore a U.S. Fish & Wildlife Service [interactive map](#) of critical habitat for endangered and threatened species in the United States. To begin, select “Visit the online mapper.”

For the most part, fishery extinction is not equivalent to biological extinction—the last fish of a species is rarely fished out of the ocean. At the same time, fishery extinction is still harmful to fish species and their ecosystems. There are some instances in which true extinction is a possibility. Whales have slow-growing populations and are at risk of complete extinction through hunting. There are some species of sharks with restricted distributions that are at risk of extinction. The groupers are another population of generally slow-growing fishes that, in the Caribbean, includes a number of species that are at risk of extinction from overfishing.

Coral reefs are extremely diverse marine ecosystems that face peril from several processes. Reefs are home to 1/3 of the world's marine fish species—about 4,000 species—despite making up only 1 percent of marine habitat. Most home marine aquaria are stocked with wild-caught organisms, not cultured organisms. Although no species is known to have been driven extinct by the pet trade in marine species, there are studies showing that populations of some species have declined in response to harvesting, indicating that the harvest is not sustainable at those levels. There are concerns about the effect of the pet trade on some terrestrial species such as turtles, amphibians, birds, plants, and even the orangutan.

**Note:**

Link to Learning



View a [brief video](#) discussing the role of marine ecosystems in supporting human welfare and the decline of ocean ecosystems.

**Bush meat** is the generic term used for wild animals killed for food. Hunting is practiced throughout the world, but hunting practices, particularly in equatorial Africa and parts of Asia, are believed to threaten several species with extinction. Traditionally, bush meat in Africa was hunted to feed families directly; however, recent commercialization of the practice now has bush meat available in grocery stores, which has increased harvest rates to the level of unsustainability. Additionally, human population growth has increased the need for protein foods that are not being met from agriculture. Species threatened by the bush meat trade are mostly mammals including many primates living in the Congo basin.

## **Exotic Species**

**Exotic species** are species that have been intentionally or unintentionally introduced by humans into an ecosystem in which they did not evolve. Such introductions likely occur frequently as natural phenomena. For example, Kudzu (*Pueraria lobata*), which is native to Japan, was introduced in the United States in 1876. It was later planted for soil conservation. Problematically, it grows too well in the southeastern United States—up to a foot a day. It is now a pest species and covers over 7 million acres in the southeastern United States. If an introduced species is able to survive in its new habitat, that introduction is now reflected in the observed range of the species. Human transportation of people and goods, including the intentional transport of organisms for trade, has dramatically increased the introduction of species into new ecosystems, sometimes at distances that are well beyond the capacity of the species to ever travel itself and outside the range of the species' natural predators.

Most exotic species introductions probably fail because of the low number of individuals introduced or poor adaptation to the ecosystem they enter. Some species, however, possess preadaptations that can make them especially successful in a new ecosystem. These exotic species often undergo dramatic population increases in their new habitat and reset the ecological conditions in the new environment, threatening the species that exist there. For this reason, exotic species are also called invasive species. Exotic species can threaten other species through competition for resources, predation, or disease.

**Note:**

## Link to Learning



Explore an [interactive global database](#) of exotic or invasive species.

Lakes and islands are particularly vulnerable to extinction threats from introduced species. In Lake Victoria, as mentioned earlier, the intentional introduction of the Nile perch was largely responsible for the extinction of about 200 species of cichlids. The accidental introduction of the brown tree snake via aircraft ([link](#)) from the Solomon Islands to Guam in 1950 has led to the extinction of three species of birds and three to five species of reptiles endemic to the island. Several other species are still threatened. The brown tree snake is adept at exploiting human transportation as a means to migrate; one was even found on an aircraft arriving in Corpus Christi, Texas. Constant vigilance on the part of airport, military, and commercial aircraft personnel is required to prevent the snake from moving from Guam to other islands in the Pacific, especially Hawaii. Islands do not make up a large area of land on the globe, but they do contain a disproportionate number of endemic species because of their isolation from mainland ancestors.



The brown tree snake, *Boiga irregularis*, is an exotic species that has caused numerous extinctions on the island of Guam since its accidental introduction in 1950. (credit: NPS)

It now appears that the global decline in amphibian species recognized in the 1990s is, in some part, caused by the fungus *Batrachochytrium dendrobatidis*, which causes the disease **chytridiomycosis** ([\[link\]](#)). There is evidence that the fungus is native to Africa and may have been spread throughout the world by transport of a commonly used laboratory and pet species: the African clawed toad (*Xenopus laevis*). It may well be that biologists themselves are responsible for spreading this disease worldwide. The North American bullfrog, *Rana catesbeiana*, which has also been widely introduced as a food animal but which easily escapes captivity, survives most infections of *Batrachochytrium dendrobatidis* and can act as a reservoir for the disease.



This Limosa Harlequin Frog (*Atelopus limosus*), an endangered species from Panama, died from a fungal disease called chytridiomycosis. The red lesions are symptomatic of the disease. (credit: Brian Gratwicke)

Early evidence suggests that another fungal pathogen, *Geomyces destructans*, introduced from Europe is responsible for **white-nose syndrome**, which infects cave-hibernating bats in eastern North America and has spread from a point of origin in western New York State ([\[link\]](#)). The disease has decimated bat populations and threatens extinction of species already listed as endangered: the Indiana bat, *Myotis sodalis*, and potentially the Virginia big-eared bat, *Corynorhinus townsendii virginianus*. How the fungus was introduced is unclear, but one logical presumption would be that recreational cavers unintentionally brought the fungus on clothes or equipment from Europe.



This little brown bat in  
Greeley Mine, Vermont,  
March 26, 2009, was found  
to have white-nose syndrome.  
(credit: Marvin Moriarty,  
USFWS)

## **Climate Change**

Climate change, and specifically the anthropogenic (meaning, caused by humans) warming trend presently underway, is recognized as a major extinction threat, particularly when combined with other threats such as habitat loss. Scientists disagree about the likely magnitude of the effects, with extinction rate estimates ranging from 15 percent to 40 percent of species committed to extinction by 2050. Scientists do agree, however, that climate change will alter regional climates, including rainfall and snowfall patterns, making habitats less hospitable to the species living in them. The warming trend will shift colder climates toward the north and south poles, forcing species to move with their adapted climate norms while facing habitat gaps along the way. The shifting ranges will impose new

competitive regimes on species as they find themselves in contact with other species not present in their historic range. One such unexpected species contact is between polar bears and grizzly bears. Previously, these two species had separate ranges. Now, their ranges are overlapping and there are documented cases of these two species mating and producing viable offspring. Changing climates also throw off species' delicate timing adaptations to seasonal food resources and breeding times. Many contemporary mismatches to shifts in resource availability and timing have already been documented.



Since 2008, grizzly bears (*Ursus arctos horribilis*) have been spotted farther north than their historic range, a possible consequence of climate change.



As a result, grizzly bear habitat now overlaps polar bear (*Ursus maritimus*) habitat. The two kinds of bears, which are capable of mating and producing viable offspring, are considered separate species as historically they lived in different habitats and never met. However, in 2006 a hunter shot a wild grizzly-polar bear hybrid known as a grolar bear, the first wild hybrid ever found.

Range shifts are already being observed: for example, some European bird species ranges have moved 91 km northward. The same study suggested that the optimal shift based on warming trends was double that distance, suggesting that the populations are not moving quickly enough. Range shifts have also been observed in plants, butterflies, other insects, freshwater fishes, reptiles, and mammals.

Climate gradients will also move up mountains, eventually crowding species higher in altitude and eliminating the habitat for those species adapted to the highest elevations. Some climates will completely disappear. The rate of warming appears to be accelerated in the arctic, which is recognized as a serious threat to polar bear populations that require sea ice to hunt seals during the winter months: seals are the only source of protein available to polar bears. A trend to decreasing sea ice coverage has occurred since observations began in the mid-twentieth century. The rate of decline observed in recent years is far greater than previously predicted by climate models.

Finally, global warming will raise ocean levels due to melt water from glaciers and the greater volume of warmer water. Shorelines will be inundated, reducing island size, which will have an effect on some species,

and a number of islands will disappear entirely. Additionally, the gradual melting and subsequent refreezing of the poles, glaciers, and higher elevation mountains—a cycle that has provided freshwater to environments for centuries—will also be jeopardized. This could result in an overabundance of salt water and a shortage of fresh water.

## **Section Summary**

The core threats to biodiversity are human population growth and unsustainable resource use. To date, the most significant causes of extinctions are habitat loss, introduction of exotic species, and overharvesting. Climate change is predicted to be a significant cause of extinctions in the coming century. Habitat loss occurs through deforestation, damming of rivers, and other activities. Overharvesting is a threat particularly to aquatic species, while the taking of bush meat in the humid tropics threatens many species in Asia, Africa, and the Americas. Exotic species have been the cause of a number of extinctions and are especially damaging to islands and lakes. Exotic species' introductions are increasing because of the increased mobility of human populations and growing global trade and transportation. Climate change is forcing range changes that may lead to extinction. It is also affecting adaptations to the timing of resource availability that negatively affects species in seasonal environments. The impacts of climate change are greatest in the arctic. Global warming will also raise sea levels, eliminating some islands and reducing the area of all others.

## **Art Connections**

### **Exercise:**

#### **Problem:**

Converting a prairie to a farm field is an example of \_\_\_\_\_.

- a. overharvesting
- b. habitat loss
- c. exotic species

d. climate change

---

**Solution:**

B

## Review Questions

**Exercise:**

**Problem:**

Which two extinction risks may be a direct result of the pet trade?

- a. climate change and exotic species introduction
- b. habitat loss and overharvesting
- c. overharvesting and exotic species introduction
- d. habitat loss and climate change

---

**Solution:**

C

**Exercise:**

**Problem:**

Exotic species are especially threatening to what kind of ecosystem?

- a. deserts
- b. marine ecosystems
- c. islands
- d. tropical forests

---

**Solution:**

## C

### Free Response

#### Exercise:

##### Problem:

Describe the mechanisms by which human population growth and resource use causes increased extinction rates.

---

##### Solution:

Human population growth leads to unsustainable resource use, which causes habitat destruction to build new human settlements, create agricultural fields, and so on. Larger human populations have also led to unsustainable fishing and hunting of wild animal populations. Excessive use of fossil fuels also leads to global warming.

#### Exercise:

##### Problem:

Explain what extinction threats a frog living on a mountainside in Costa Rica might face.

---

##### Solution:

The frog is at risk from global warming shifting its preferred habitat up the mountain. In addition, it will be at risk from exotic species, either as a new predator or through the impact of transmitted diseases such as chytridiomycosis. It is also possible that habitat destruction will threaten the species.

### Glossary

bush meat

wild-caught animal used as food (typically mammals, birds, and reptiles); usually referring to hunting in the tropics of sub-Saharan Africa, Asia, and the Americas

chytridiomycosis

disease of amphibians caused by the fungus *Batrachochytrium dendrobatidis*; thought to be a major cause of the global amphibian decline

exotic species

(also, invasive species) species that has been introduced to an ecosystem in which it did not evolve

tragedy of the commons

economic principle that resources held in common will inevitably be overexploited

white-nose syndrome

disease of cave-hibernating bats in the eastern United States and Canada associated with the fungus *Geomyces destructans*

## Preserving Biodiversity

By the end of this section, you will be able to:

- Identify new technologies for describing biodiversity
- Explain the legislative framework for conservation
- Describe principles and challenges of conservation preserve design
- Identify examples of the effects of habitat restoration
- Discuss the role of zoos in biodiversity conservation

Preserving biodiversity is an extraordinary challenge that must be met by greater understanding of biodiversity itself, changes in human behavior and beliefs, and various preservation strategies.

## Measuring Biodiversity

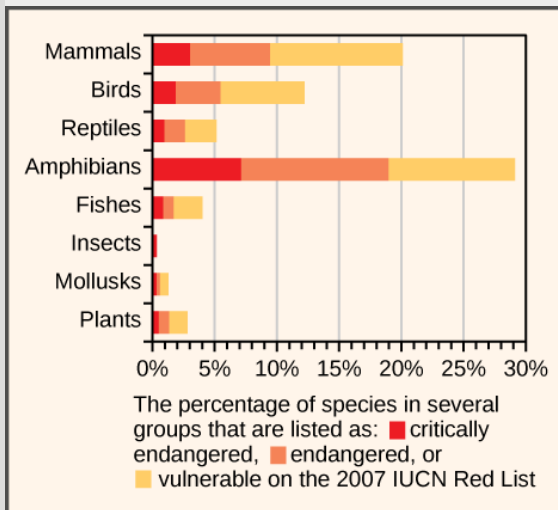
The technology of molecular genetics and data processing and storage are maturing to the point where cataloguing the planet's species in an accessible way is close to feasible. **DNA barcoding** is one molecular genetic method, which takes advantage of rapid evolution in a mitochondrial gene present in eukaryotes, excepting the plants, to identify species using the sequence of portions of the gene. Plants may be barcoded using a combination of chloroplast genes. Rapid mass sequencing machines make the molecular genetics portion of the work relatively inexpensive and quick. Computer resources store and make available the large volumes of data. Projects are currently underway to use DNA barcoding to catalog museum specimens, which have already been named and studied, as well as testing the method on less studied groups. As of mid 2012, close to 150,000 named species had been barcoded. Early studies suggest there are significant numbers of undescribed species that looked too much like sibling species to previously be recognized as different. These now can be identified with DNA barcoding.

Numerous computer databases now provide information about named species and a framework for adding new species. However, as already noted, at the present rate of description of new species, it will take close to 500 years before the complete catalog of life is known. Many, perhaps most, species on the planet do not have that much time.

There is also the problem of understanding which species known to science are threatened and to what degree they are threatened. This task is carried out by the non-profit IUCN which, as previously mentioned, maintains the Red List—an online listing of endangered species categorized by taxonomy, type of threat, and other criteria ([link](#)). The Red List is supported by scientific research. In 2011, the list contained 61,000 species, all with supporting documentation.

**Note:**

**Art Connection**



This chart shows the percentage of various animal species, by group, on the IUCN Red List as of 2007.

Which of the following statements is not supported by this graph?

- There are more vulnerable fishes than critically endangered and endangered fishes combined.
- There are more critically endangered amphibians than vulnerable, endangered and critically endangered reptiles combined.

- c. Within each group, there are more critically endangered species than vulnerable species.
- d. A greater percentage of bird species are critically endangered than mollusk species.

## **Changing Human Behavior**

Legislation throughout the world has been enacted to protect species. The legislation includes international treaties as well as national and state laws. The Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES) treaty came into force in 1975. The treaty, and the national legislation that supports it, provides a legal framework for preventing approximately 33,000 listed species from being transported across nations' borders, thus protecting them from being caught or killed when international trade is involved. The treaty is limited in its reach because it only deals with international movement of organisms or their parts. It is also limited by various countries' ability or willingness to enforce the treaty and supporting legislation. The illegal trade in organisms and their parts is probably a market in the hundreds of millions of dollars. Illegal wildlife trade is monitored by another non-profit: Trade Records Analysis of Flora and Fauna in Commerce (TRAFFIC).

Within many countries there are laws that protect endangered species and regulate hunting and fishing. In the United States, the Endangered Species Act (ESA) was enacted in 1973. Species at risk are listed by the Act; the U.S. Fish & Wildlife Service is required by law to develop management plans that protect the listed species and bring them back to sustainable numbers. The Act, and others like it in other countries, is a useful tool, but it suffers because it is often difficult to get a species listed, or to get an effective management plan in place once it is listed. Additionally, species may be controversially taken off the list without necessarily having had a change in their situation. More fundamentally, the approach to protecting individual species rather than entire ecosystems is both inefficient and focuses efforts on a few highly visible and often charismatic species, perhaps at the expense of other species that go unprotected. At the same



time, the Act has a critical habitat provision outlined in the recovery mechanism that may benefit species other than the one targeted for management.

The Migratory Bird Treaty Act (MBTA) is an agreement between the United States and Canada that was signed into law in 1918 in response to declines in North American bird species caused by hunting. The Act now lists over 800 protected species. It makes it illegal to disturb or kill the protected species or distribute their parts (much of the hunting of birds in the past was for their feathers).

The international response to global warming has been mixed. The Kyoto Protocol, an international agreement that came out of the United Nations Framework Convention on Climate Change that committed countries to reducing greenhouse gas emissions by 2012, was ratified by some countries, but spurned by others. Two important countries in terms of their potential impact that did not ratify the Kyoto Protocol were the United States and China. The United States rejected it as a result of a powerful fossil fuel industry and China because of a concern it would stifle the nation's growth. Some goals for reduction in greenhouse gasses were met and exceeded by individual countries, but worldwide, the effort to limit greenhouse gas production is not succeeding. The intended replacement for the Kyoto Protocol has not materialized because governments cannot agree on timelines and benchmarks. Meanwhile, climate scientists predict the resulting costs to human societies and biodiversity will be high.

As already mentioned, the private non-profit sector plays a large role in the conservation effort both in North America and around the world. The approaches range from species-specific organizations to the broadly focused IUCN and TRAFFIC. The Nature Conservancy takes a novel approach. It purchases land and protects it in an attempt to set up preserves for ecosystems. Ultimately, human behavior will change when human values change. At present, the growing urbanization of the human population is a force that poses challenges to the valuing of biodiversity.

## **Conservation in Preserves**

Establishment of wildlife and ecosystem preserves is one of the key tools in conservation efforts. A preserve is an area of land set aside with varying degrees of protection for the organisms that exist within the boundaries of the preserve. Preserves can be effective in the short term for protecting both species and ecosystems, but they face challenges that scientists are still exploring to strengthen their viability as long-term solutions.

### **How Much Area to Preserve?**

Due to the way protected lands are allocated (they tend to contain less economically valuable resources rather than being set aside specifically for the species or ecosystems at risk) and the way biodiversity is distributed, determining a target percentage of land or marine habitat that should be protected to maintain biodiversity levels is challenging. The IUCN World Parks Congress estimated that 11.5 percent of Earth's land surface was covered by preserves of various kinds in 2003. This area is greater than previous goals; however, it only represents 9 out of 14 recognized major biomes. Research has shown that 12 percent of all species live only outside preserves; these percentages are much higher when only threatened species and high quality preserves are considered. For example, high quality preserves include only about 50 percent of threatened amphibian species. The conclusion must be that either the percentage of area protected must increase, or the percentage of high quality preserves must increase, or preserves must be targeted with greater attention to biodiversity protection. Researchers argue that more attention to the latter solution is required.

### **Preserve Design**

There has been extensive research into optimal preserve designs for maintaining biodiversity. The fundamental principle behind much of the research has been the seminal theoretical work of Robert H. MacArthur and Edward O. Wilson published in 1967 on island biogeography.<sup>[\[footnote\]](#)</sup> This work sought to understand the factors affecting biodiversity on islands. The fundamental conclusion was that biodiversity on an island was a function of

the origin of species through migration, speciation, and extinction on that island. Islands farther from a mainland are harder to get to, so migration is lower and the equilibrium number of species is lower. Within island populations, evidence suggests that the number of species gradually increases to a level similar to the numbers on the mainland from which the species is suspected to have migrated. In addition, smaller islands are harder to find, so their immigration rates for new species are lower. Smaller islands are also less geographically diverse so there are fewer niches to promote speciation. And finally, smaller islands support smaller populations, so the probability of extinction is higher.

Robert H. MacArthur and Edward O. Wilson, E. O., *The Theory of Island Biogeography* (Princeton, N.J.: Princeton University Press, 1967).

As islands get larger, the number of species accelerates, although the effect of island area on species numbers is not a direct correlation. Conservation preserves can be seen as “islands” of habitat within “an ocean” of non-habitat. For a species to persist in a preserve, the preserve must be large enough. The critical size depends, in part, on the home range that is characteristic of the species. A preserve for wolves, which range hundreds of kilometers, must be much larger than a preserve for butterflies, which might range within ten kilometers during its lifetime. But larger preserves have more core area of optimal habitat for individual species, they have more niches to support more species, and they attract more species because they can be found and reached more easily.

Preserves perform better when there are buffer zones around them of suboptimal habitat. The buffer allows organisms to exit the boundaries of the preserve without immediate negative consequences from predation or lack of resources. One large preserve is better than the same area of several smaller preserves because there is more core habitat unaffected by edges. For this same reason, preserves in the shape of a square or circle will be better than a preserve with many thin “arms.” If preserves must be smaller, then providing wildlife corridors between them so that individuals and their genes can move between the preserves, for example along rivers and streams, will make the smaller preserves behave more like a large one. All of these factors are taken into consideration when planning the nature of a preserve before the land is set aside.

In addition to the physical, biological, and ecological specifications of a preserve, there are a variety of policy, legislative, and enforcement specifications related to uses of the preserve for functions other than protection of species. These can include anything from timber extraction, mineral extraction, regulated hunting, human habitation, and nondestructive human recreation. Many of these policy decisions are made based on political pressures rather than conservation considerations. In some cases, wildlife protection policies have been so strict that subsistence-living indigenous populations have been forced from ancestral lands that fell within a preserve. In other cases, even if a preserve is designed to protect wildlife, if the protections are not or cannot be enforced, the preserve status will have little meaning in the face of illegal poaching and timber extraction. This is a widespread problem with preserves in areas of the tropics.

### **Limitations on Preserves**

Some of the limitations on preserves as conservation tools are evident from the discussion of preserve design. Political and economic pressures typically make preserves smaller, never larger, so setting aside areas that are large enough is difficult. If the area set aside is sufficiently large, there may not be sufficient area to create a buffer around the preserve. In this case, an area on the outer edges of the preserve inevitably becomes a riskier suboptimal habitat for the species in the preserve. Enforcement of protections is also a significant issue in countries without the resources or political will to prevent poaching and illegal resource extraction.

Climate change will create inevitable problems with the location of preserves. The species within them will migrate to higher latitudes as the habitat of the preserve becomes less favorable. Scientists are planning for the effects of global warming on future preserves and striving to predict the need for new preserves to accommodate anticipated changes to habitats; however, the end effectiveness is tenuous since these efforts are prediction based.

Finally, an argument can be made that conservation preserves reinforce the cultural perception that humans are separate from nature, can exist outside of it, and can only operate in ways that do damage to biodiversity. Creating preserves reduces the pressure on human activities outside the preserves to be sustainable and non-damaging to biodiversity. Ultimately, the political, economic, and human demographic pressures will degrade and reduce the size of conservation preserves if the activities outside them are not altered to be less damaging to biodiversity.

**Note:**

**Link to Learning**



An [interactive global data system](#) of protected areas can be found at website. Review data about individual protected areas by location or study statistics on protected areas by country or region.

## **Habitat Restoration**

Habitat restoration holds considerable promise as a mechanism for restoring and maintaining biodiversity. Of course once a species has become extinct, its restoration is impossible. However, restoration can improve the biodiversity of degraded ecosystems. Reintroducing wolves, a top predator, to Yellowstone National Park in 1995 led to dramatic changes in the ecosystem that increased biodiversity. The wolves ([link](#)) function to suppress elk and coyote populations and provide more abundant resources to the guild of carrion eaters. Reducing elk populations has allowed revegetation of riparian areas, which has increased the diversity of species in that habitat. Decreasing the coyote population has increased the

populations of species that were previously suppressed by this predator. The number of species of carrion eaters has increased because of the predatory activities of the wolves. In this habitat, the wolf is a keystone species, meaning a species that is instrumental in maintaining diversity in an ecosystem. Removing a keystone species from an ecological community may cause a collapse in diversity. The results from the Yellowstone experiment suggest that restoring a keystone species can have the effect of restoring biodiversity in the community. Ecologists have argued for the identification of keystone species where possible and for focusing protection efforts on those species; likewise, it also makes sense to attempt to return them to their ecosystem if they have been removed.



(a)



(b)



(c)



(d)

(a) The Gibbon wolf pack in Yellowstone National Park, March 1, 2007, represents a keystone species. The reintroduction of wolves into Yellowstone National Park in 1995 led to a change in the grazing behavior of (b) elk. To avoid predation, the elk no longer grazed exposed stream and riverbeds, such as (c) the Lamar Riverbed in Yellowstone. This allowed willow and cottonwood seedlings to grow. The

seedlings decreased erosion and provided shading to the creek, which improved fish habitat. A new colony of (d) beaver may also have benefited from the habitat change. (credit a: modification of work by Doug Smith, NPS; credit c: modification of work by Jim Peaco, NPS; credit d: modification of work by “Shiny Things”/Flickr)

Other large-scale restoration experiments underway involve dam removal. In the United States, since the mid-1980s, many aging dams are being considered for removal rather than replacement because of shifting beliefs about the ecological value of free-flowing rivers and because many dams no longer provide the benefit and functions that they did when they were first built. The measured benefits of dam removal include restoration of naturally fluctuating water levels (the purpose of dams is frequently to reduce variation in river flows), which leads to increased fish diversity and improved water quality. In the Pacific Northwest, dam removal projects are expected to increase populations of salmon, which is considered a keystone species because it transports key nutrients to inland ecosystems during its annual spawning migrations. In other regions such as the Atlantic coast, dam removal has allowed the return of spawning anadromous fish species (species that are born in fresh water, live most of their lives in salt water, and return to fresh water to spawn). Some of the largest dam removal projects have yet to occur or have happened too recently for the consequences to be measured. The large-scale ecological experiments that these removal projects constitute will provide valuable data for other dam projects slated either for removal or construction.

## **The Role of Captive Breeding**

Zoos have sought to play a role in conservation efforts both through captive breeding programs and education. The transformation of the missions of zoos from collection and exhibition facilities to organizations that are

dedicated to conservation is ongoing. In general, it has been recognized that, except in some specific targeted cases, captive breeding programs for endangered species are inefficient and often prone to failure when the species are reintroduced to the wild. Zoo facilities are far too limited to contemplate captive breeding programs for the numbers of species that are now at risk. Education is another potential positive impact of zoos on conservation efforts, particularly given the global trend to urbanization and the consequent reduction in contacts between people and wildlife. A number of studies have been performed to look at the effectiveness of zoos on people's attitudes and actions regarding conservation; at present, the results tend to be mixed.

## **Section Summary**

New technological methods such as DNA barcoding and information processing and accessibility are facilitating the cataloging of the planet's biodiversity. There is also a legislative framework for biodiversity protection. International treaties such as CITES regulate the transportation of endangered species across international borders. Legislation within individual countries protecting species and agreements on global warming have had limited success; there is at present no international agreement on targets for greenhouse gas emissions. In the United States, the Endangered Species Act protects listed species but is hampered by procedural difficulties and a focus on individual species. The Migratory Bird Act is an agreement between Canada and the United States to protect migratory birds. The non-profit sector is also very active in conservation efforts in a variety of ways.

Conservation preserves are a major tool in biodiversity protection. Presently, 11percent of Earth's land surface is protected in some way. The science of island biogeography has informed the optimal design of preserves; however, preserves have limitations imposed by political and economic forces. In addition, climate change will limit the effectiveness of preserves in the future. A downside of preserves is that they may lessen the pressure on human societies to function more sustainably outside the preserves.



Habitat restoration has the potential to restore ecosystems to previous biodiversity levels before species become extinct. Examples of restoration include reintroduction of keystone species and removal of dams on rivers. Zoos have attempted to take a more active role in conservation and can have a limited role in captive breeding programs. Zoos also may have a useful role in education.

## Art Connections

### Exercise:

#### Problem:

[\[link\]](#) Which of the following statements is not supported by this graph?

- a. There are more vulnerable fishes than critically endangered and endangered fishes combined.
- b. There are more critically endangered amphibians than vulnerable, endangered and critically endangered reptiles combined.
- c. Within each group, there are more critically endangered species than vulnerable species.
- d. A greater percentage of bird species are critically endangered than mollusk species.

---

#### Solution:

[\[link\]](#) C

## Review Questions

### Exercise:

**Problem:**

Certain parrot species cannot be brought to the United States to be sold as pets. What is the name of the legislation that makes this illegal?

- a. Red List
- b. Migratory Bird Act
- c. CITES
- d. Endangered Species Act (ESA)

---

**Solution:**

C

**Exercise:****Problem:**

What was the name of the first international agreement on climate change?

- a. Red List
- b. Montreal Protocol
- c. International Union for the Conservation of Nature (IUCN)
- d. Kyoto Protocol

---

**Solution:**

D

**Exercise:****Problem:**

About what percentage of land on the planet is set aside as a preserve of some type?

- a. 1 percent

- b. 6 percent
- c. 11 percent
- d. 15 percent

---

**Solution:**

C

## Free Response

**Exercise:**

**Problem:** Describe two considerations in conservation preserve design.

---

**Solution:**

Larger preserves will contain more species. Preserves should have a buffer around them to protect species from edge effects. Preserves that are round or square are better than preserves with many thin arms.

**Exercise:**

**Problem:**

Describe what happens to an ecosystem when a keystone species is removed.

---

**Solution:**











When a keystone species is removed many species will disappear from the ecosystem.

## Glossary

DNA barcoding

molecular genetic method for identifying a unique genetic sequence to associate with a species

<div>Group</div>		Periodic Table of the Elements																18																			
1	H 1.01 Hydrogen																	2	He 4.00 Helium																		
2	Li 6.94 Lithium	3	Be 9.01 Beryllium															10	Ne 20.18 Neon																		
3	B 10.81 Boron	4	C 12.11 Carbon	5	N 14.01 Nitrogen	6	O 15.99 Oxygen	7	F 18.99 Fluorine	8	Ar 39.95 Argon																										
11	Na 22.99 Sodium	12	Mg 24.31 Magnesium									13	Al 26.98 Aluminum	14	Si 28.09 Silicon	15	P 30.97 Phosphorus	16	S 32.07 Sulfur	17	Cl 35.45 Chlorine	18	Ar 39.95 Argon														
19	K 39.09 Potassium	20	Ca 40.08 Calcium	21	Sc 44.96 Scandium	22	Ti 47.87 Titanium	23	V 50.94 Vanadium	24	Cr 51.99 Chromium	25	Mn 54.94 Manganese	26	Fe 55.85 Iron	27	Co 58.93 Cobalt	28	Ni 58.69 Nickel	29	Cu 63.55 Copper	30	Zn 65.41 Zinc	31	Ga 69.72 Gallium	32	Ge 72.64 Germanium	33	As 74.92 Arsenic	34	Se 78.96 Selenium	35	Br 79.90 Bromine	36	Kr 83.79 Krypton		
37	Rb 85.47 Rubidium	38	Sr 87.62 Strontium	39	Y 88.91 Yttrium	40	Zr 91.22 Zirconium	41	Nb 92.91 Niobium	42	Mo 95.94 Molybdenum	43	Tc [98] Technetium	44	Ru 101.1 Ruthenium	45	Rh 102.9 Rhodium	46	Pd 106.4 Palladium	47	Ag 107.9 Silver	48	Cd 112.4 Cadmium	49	In 114.8 Indium	50	Sn 118.7 Tin	51	Sb 121.8 Antimony	52	Te 127.6 Tellurium	53	I 126.9 Iodine	54	Xe 131.3 Xenon		
55	Cs 132.9 Cesium	56	Ba 137.3 Barium	* <td>57-71 <b>Lanthanides</b></td> <td>**<td>72-103 <b>Actinides</b></td><td>73</td><td>Hf 178.5 Hafnium</td><td>74</td><td>Ta 180.9 Tantalum</td><td>75</td><td>W 183.8 Tungsten</td><td>76</td><td>Re 186.2 Rhenium</td><td>77</td><td>Os 190.2 Osmium</td><td>78</td><td>Ir 192.2 Iridium</td><td>79</td><td>Pt 195.1 Platinum</td><td>80</td><td>Au 196.9 Gold</td><td>81</td><td>Hg 200.6 Mercury</td><td>82</td><td>Tl 204.4 Thallium</td><td>83</td><td>Pb 207.2 Lead</td><td>84</td><td>Bi 208.9 Bismuth</td><td>85</td><td>Po [209] Polonium</td><td>86</td><td>At [210] Astatine</td><td>87</td><td>Rn [222] Radon</td></td>	57-71 <b>Lanthanides</b>	** <td>72-103 <b>Actinides</b></td> <td>73</td> <td>Hf 178.5 Hafnium</td> <td>74</td> <td>Ta 180.9 Tantalum</td> <td>75</td> <td>W 183.8 Tungsten</td> <td>76</td> <td>Re 186.2 Rhenium</td> <td>77</td> <td>Os 190.2 Osmium</td> <td>78</td> <td>Ir 192.2 Iridium</td> <td>79</td> <td>Pt 195.1 Platinum</td> <td>80</td> <td>Au 196.9 Gold</td> <td>81</td> <td>Hg 200.6 Mercury</td> <td>82</td> <td>Tl 204.4 Thallium</td> <td>83</td> <td>Pb 207.2 Lead</td> <td>84</td> <td>Bi 208.9 Bismuth</td> <td>85</td> <td>Po [209] Polonium</td> <td>86</td> <td>At [210] Astatine</td> <td>87</td> <td>Rn [222] Radon</td>	72-103 <b>Actinides</b>	73	Hf 178.5 Hafnium	74	Ta 180.9 Tantalum	75	W 183.8 Tungsten	76	Re 186.2 Rhenium	77	Os 190.2 Osmium	78	Ir 192.2 Iridium	79	Pt 195.1 Platinum	80	Au 196.9 Gold	81	Hg 200.6 Mercury	82	Tl 204.4 Thallium	83	Pb 207.2 Lead	84	Bi 208.9 Bismuth	85	Po [209] Polonium	86	At [210] Astatine	87	Rn [222] Radon
87	Fr [223] Francium	88	Ra [226] Radium	*	89-103 <b>Ac-Lu</b>	104	Rf [261] Rutherfordium	105	Db [262] Dubnium	106	Sg [266] Seaborgium	107	Bh [264] Bohrium	108	Hs [277] Hassium	109	Mt [268] Meitnerium	110	Ds [269] Darmstadtium	111	Rg [272] Roentgenium	112	Cn [285] Copernicium	113	Uut [284] Ununtrium	114	Fl [289] Flerovium	115	Uup [288] Ununpentium	116	Lv [293] Livermorium	117	Uus [294] Ununseptium	118	Uuo [294] Ununoctium		
		57	La 138.9 Lanthanum	58	Ce 140.1 Cerium	59	Pr 140.9 Praseodymium	60	Ba 144.2 Europium	61	Pm [145] Promethium	62	Sm 150.4 Samarium	63	Eu 151.9 Europium	64	Gd 157.3 Gadolinium	65	Tb 158.9 Terbium	66	Dy 162.5 Dysprosium	67	Ho 164.9 Holmium	68	Er 167.3 Erbium	69	Tm 168.9 Thulium	70	Yb 173.1 Ytterbium	71	Lu 174.9 Lutetium						
		89	Ac [227] Actinium	90	Th 232.0 Thorium	91	Pa 231.0 Protactinium	92	U 238.0 Uranium	93	Np [237] Neptunium	94	Pu [244] Plutonium	95	Am [243] Americium	96	Cm [247] Curium	97	Bk [247] Berkeleyum	98	Cf [251] Californium	99	Es [252] Einsteinium	100	Fm [257] Fermium	101	Md [258] Mendelevium	102	No [259] Nobelium	103	Lr [262] Lawrencium						

	Other non-metals		Noble gases
	Alkali metals		Lanthanides
	Transition metals		Actinides
	Other metals		Unknown chemical properties
	Alkaline earth metals		
	Halogens		